

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2003, 13:48:16 ; Search time 42 Seconds
(without alignments)
1152.657 Million cell updates/sec

Title: US-09-993-292A-2

Perfect score: 1515

Sequence: 1 MFSIFAEQTVKSAIETA.....NEYQORHGKTLFEVDPVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database: A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1515	100.0	305	24	ABG73265 Salmonella typhi c
2	1515	100.0	749	24	ABG73266 Salmonella typhi C
3	1390	91.7	309	16	ABG67645 Haemolysin regulat
4	1079	71.2	1035	22	ABG25221 Novel human diagno
5	118.5	7.8	2681	24	ABJ19025 Pathogen specific
6	114.5	7.6	588	22	ABE71280 Drosophila melanog
7	112.5	7.4	1948	23	ABF73774 Candida albicans e
8	110.5	7.3	476	14	AAE43563 Hyaluronan recepto
9	110	7.3	951	22	ABU53070 Intracellular traf

10	110	7.3	953	22	ABU53071 Intracellular traf
11	110	7.3	961	22	ABU53077 Protein differenti
12	110	7.3	2230	24	ABU07445 Putative sensory t
13	109	7.2	733	22	ABE96493 Intracellular traf
14	109	7.2	928	22	ABU53079 Drosophila melanog
15	108.5	7.2	1048	22	ABE59245 Amino acid sequenc
16	108.5	7.2	1086	22	ABG67418 Staphylococcus epi
17	108.5	7.2	1086	24	ABG72693 Fruitfly KLP61, a
18	108.5	7.2	3696	23	ABF40235 Amino acid sequenc
19	107.5	7.1	1292	23	ABE77986 Staphylococcus epi
20	107	7.1	1211	23	ABP39975 Listeria monocytog
21	106.5	7.0	927	23	ABE49720 Human RAD50. Homo
22	106.5	7.0	1312	18	AAW22775 Human homologue of
23	106.5	7.0	1312	19	AAW71295 Drosophila melanog
24	106.5	7.0	2346	22	ABE63519 Drosophila melanog
25	106.5	7.0	7201	22	ABE71136 Drosophila melanog
26	106	7.0	685	22	ABE64018 S. epidermidis ope
27	106	7.0	885	22	AAE82283 Novel signal pathw
28	106	7.0	2835	23	AAU75100 Amino acid sequenc
29	105.5	7.0	1318	23	ABE77985 N. gonorrhoeae ami
30	105	6.9	1015	24	ABF76825 N. gonorrhoeae ami
31	105	6.9	1015	24	ABF76817 Streptococcus poly
32	104.5	6.9	318	23	ABP30285 Streptococcus poly
33	104.5	6.9	339	23	ABP28740 Drosophila melanog
34	104.5	6.9	1377	22	ABE65439 Staphylococcus aur
35	104.5	6.9	3158	22	AAU37018 Lactococcus lactis
36	104	6.9	901	23	ABE55018 Human kinase, MEK1
37	104	6.9	2334	23	ABG31849 Human novel protei
38	103.5	6.8	1959	23	ABU10604 Drosophila melanog
39	103.5	6.8	2056	22	ABE59344 Human NuMA protein
40	103.5	6.8	2101	21	AAV49936 Receptor for hyalu
41	103	6.8	606	17	AAE99675 RHAMM 1-2a isoform
42	103	6.8	631	17	AAE99675 Herbicidally activ
43	103	6.8	796	23	ABE92030 An Escherichia col
44	102.5	6.8	974	21	AAE93246 Human polypeptide
45	102.5	6.8	2115	21	AAV49937 Human polypeptide
46	102	6.7	1038	23	ABF79399 Streptococcus uber
47	102	6.7	1453	22	AAE39213 Clostridium diff
48	102	6.7	1469	22	AAE39214 Mouse receptor for
49	102	6.7	1976	24	ABU07447 Mouse hyaluronon a
50	102	6.7	1988	22	AAE40399 Novel human diagno
51	102	6.7	1988	22	AAE40399 Human alpha-myosin
52	101.5	6.7	561	19	ABG63043 Human novel secret
53	101.5	6.7	610	23	ABG80347 Intracellular traf
54	101	6.7	631	23	ABG60843 B. burgdorferi ant
55	101	6.7	794	23	AAU11437 Neisseria meningit
56	101	6.7	795	22	ABG15416 Neisseria meningit
57	101	6.7	1939	23	ABE77096 N. meningitidis st
58	100.5	6.6	486	23	ABG79834 Enterococcus faeca
59	100.5	6.6	977	22	AAE25384 Human NZMS-8 prote
60	100	6.6	210	20	AAE25384 Lactococcus lactis
61	100	6.6	590	22	AAE25384 Breast cancer asso
62	100	6.6	590	22	AAE25384 Human protein SEQ
63	100	6.6	612	22	ABG01716 Novel human diagno
64	100	6.6	612	22	ABG01716 Protein differenti
65	100	6.6	722	22	AAU35203 Novel human diagno
66	100	6.6	885	23	ABE25384 Human NZMS-8 prote
67	100	6.6	1046	23	ABE25384 Lactococcus lactis
68	100	6.6	1752	20	AAE25384 Breast cancer asso
69	100	6.6	1960	22	AAE25384 Human protein SEQ
70	100	6.6	2143	22	ABG01716 Novel human diagno
71	100	6.6	2871	24	ABU07402 Protein differenti
72	100	6.6	2918	22	ABG27218 Novel human diagno
73	100	6.6	10182	23	ABF38314 Staphylococcus epi
74	99.5	6.6	630	18	AAE39166 Mouse RHAMM protei
75	99.5	6.6	660	22	AAU38921 C. trachomatis CT8
76	99.5	6.6	729	22	AAU60877 Propionibacterium
77	99.5	6.6	1259	23	ABP73283 Candida albicans e
78	99.5	6.6	1963	23	AAE79838 Human protein SEQ
79	99.5	6.6	5373	22	AAU14603 Novel bone marrow
80	99	6.5	672	22	ABE93091 Human protein sequ
81	98.5	6.5	1010	23	ABE40333 Staphylococcus epi
82	98.5	6.5	5447	22	AAU14697 Novel bone marrow

83 98 6.5 366 21 AAG13859
 84 98 6.5 366 21 AAG23551
 85 98 6.5 373 21 AAG13858
 86 98 6.5 373 21 AAG23550
 87 98 6.5 399 21 AAG13857
 88 98 6.5 401 21 AAG23549
 89 98 6.5 462 21 ABU11780
 90 98 6.5 953 22 ABU53069
 91 98 6.5 1095 24 ABU25381
 92 98 6.5 1223 23 ABP69358
 93 98 6.5 1223 23 AAU96323
 94 98 6.5 1274 23 AAU22909
 95 98 6.5 1277 24 ABU25981
 96 98 6.5 1851 22 AEG01723
 97 98 6.5 1886 19 AAW54241
 98 98 6.5 4820 22 ABB58592
 99 97.5 6.4 759 22 AAU33564
 100 97.5 6.4 1857 23 AAU84350

ALIGNMENTS

RESULT 1
 ABG73265
 ID ABG73265 standard; Protein; 305 AA.

XX AC ABG73265;
 XX DT 17-APR-2003 (first entry)
 XX DE Salmonella typhi cytolysin A (ClyA) protein.
 XX KW Protein expression vector; expression cassette; export protein;
 KW transformed host cell; bacterial cell; immune response;
 KW HlyE family; cytolysin A; ClyA; recombinant protein; bioreactor;
 KW bacterial growth; live vector immunogenic composition; animal feed;
 KW animal vaccination; food industry; nutritional supplement; biomediation;
 KW waste disposal; waste treatment; host bacterium.
 XX OS Salmonella typhi.
 XX FN US2002146430-A1.
 XX PD 10-OCT-2002.
 XX PF 23-NOV-2001; 2001US-0993292.
 XX PR 22-NOV-2000; 2000US-252516P.
 XX PA (GALE/) GALEN J E.
 XX PI Galen JE;
 XX DR WPI: 2003-228013/22.
 XX DR N-FSDB; ABX15174.
 XX PT Novel system for expressing protein of interest, has expression vector
 PT with cassette having export protein coding sequence fused to protein
 PT coding sequence, host cell transformed with vector, and culturing
 PT environment -
 XX PS Claim 5; Page 18; 33pp; English.
 XX CC The present invention relates to a system for expressing a protein
 CC of interest. The system comprises an expression vector having an
 CC export protein cassette, where the expression cassette comprises an
 CC export protein coding sequence genetically fused to a DNA sequence
 CC encoding the protein of interest, a host cell transformed with
 CC the expression vector, and a culturing environment for the transformed
 CC host cell, where the expression cassette expresses the export
 CC protein-protein of interest fusion protein which is exported out of
 CC the host cell. The system is useful for expressing a gene in a

CC bacterial cell, by providing an expression vector to a population of
 CC untransformed bacterial cells, and expressing the expression cassette,
 CC such that the export protein-protein of interest fusion protein is
 CC produced and exported into the culture medium. The protein of
 CC interest is preferably an antigen and this method is useful for
 CC eliciting an immune response from a host against the fusion protein.
 CC The export protein may be a member of the HlyE family such as
 CC cytolysin A (ClyA). The system is useful for the production of
 CC recombinant proteins inside a bacterial host cell, in a bioreactor,
 CC and various devices that facilitate the growth of bacteria. The system
 CC is also useful for providing an animal antigenic material against which
 CC an immune response may be mounted. The obtained recombinant bacterium
 CC (e.g. Salmonella) can be used as a live vector immunogenic composition
 CC capable of facilitating the generation of an immune response
 CC in an animal. The system is also useful for preparing immunogenic
 CC compositions for vaccinating animals, and for use in the food,
 CC nutritional supplement, animal feed, biomediation, waste disposal,
 CC and waste treatment industries. The system is very advantageous for
 CC protein expression, as proteins that are toxic to host bacterium can
 CC also be expressed. A population of recombinant host cells can be
 CC transformed with the expression vector, and then the population of
 CC bacterial host cells can be maintained in culture and used to produce
 CC proteins without having to harvest and lyse the host cells. The present
 CC represents Salmonella typhi ClyA protein.

XX Sequence 305 AA;

Query Match 100.0%; Score 1515; DB 24; Length 305;
 Best Local Similarity 100.0%; Pred. No. 3.4e-116;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTSIFAQTVBVVKSATETADGALDLYNKYLDQVIMKTFDETIKELSRFKQESQASV 60
 DB 1 MTSIFAQTVBVVKSATETADGALDLYNKYLDQVIMKTFDETIKELSRFKQESQASV 60
 QY 61 LVGDIKVLMLDSQKYFEATQTVYEWGCVTOLLISAVILLPDEYNEKKAQAQKILIRIL 120
 DB 61 LVGDIKVLMLDSQKYFEATQTVYEWGCVTOLLISAVILLPDEYNEKKAQAQKILIRIL 120
 QY 121 DGVKVLNEAKSKLLTSSQSFNNASGKLALDLSOLTDFSEKSYFOSQVDRIRKEAYAG 180
 DB 121 DGVKVLNEAKSKLLTSSQSFNNASGKLALDLSOLTDFSEKSYFOSQVDRIRKEAYAG 180
 QY 181 AAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNELKTQVNPFTSLSTVQANKDIDAAK 240
 DB 181 AAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNELKTQVNPFTSLSTVQANKDIDAAK 240
 QY 241 LKLAETIAAIGEIKTETETTRFYVDYDDLMLSLKGAAKWINTCNEYQQRHGKKTLEFV 300
 DB 241 LKLAETIAAIGEIKTETETTRFYVDYDDLMLSLKGAAKWINTCNEYQQRHGKKTLEFV 300
 QY 301 PDVAS 305
 DB 301 PDVAS 305

RESULT 2

ABG73266
 ID ABG73266 standard; Protein; 749 AA.
 XX AC ABG73266;
 XX DT 17-APR-2003 (first entry)
 XX DE Salmonella typhi ClyA-Bacillus subtilis sacB fusion protein.

XX KW Protein expression vector; expression cassette; export protein;
 KW transformed host cell; bacterial cell; immune response;
 KW HlyE family; cytolysin A; ClyA; recombinant protein; bioreactor;
 KW bacterial growth; live vector immunogenic composition; animal feed;
 KW animal vaccination; food industry; nutritional supplement; biomediation;
 KW waste disposal; waste treatment; host bacterium; sacB; mutant; mutagen;
 KW lethal levansucrase.

XX OS Chimeric - Salmonella typhi.
 OS Chimeric - Bacillus subtilis.
 OS Synthetic.
 XX PN US2002146430-A1.
 XX PD 10-OCT-2002.
 XX PF 23-NOV-2001; 2001US-09933292.
 XX PR 22-NOV-2000; 2000US-452516P.
 XX PA (GALE/J) GALEN J E.
 XX PI Galen JE;
 XX DR WPI: 2003-228013/22.
 XX DR N-PSDB; ABX15191.
 XX PT Novel system for expressing protein of interest, has expression vector
 PT with cassette having export protein coding sequence fused to protein
 PT coding sequence, host cell transformed with vector, and culturing
 PT environment -
 XX PS Example 3; Page 25-28; 33pp; English.
 XX The present invention relates to a system for expressing a protein
 CC of interest. The system comprises an expression vector having an
 CC expression cassette, where the expression cassette comprises an
 CC export protein coding sequence genetically fused to a DNA sequence
 CC encoding the protein of interest, a host cell transformed with
 CC the expression vector, and a culturing environment for the transformed
 CC host cell, where the expression cassette expresses the export
 CC protein-protein of interest fusion protein which is exported out of
 CC the host cell. The system is useful for expressing a gene in a
 CC bacterial cell, by providing an expression vector to a population of
 CC untransformed bacterial cells, and expressing the expression cassette,
 CC such that the export protein-protein of interest fusion protein is
 CC produced and exported into the culture medium. The protein of
 CC interest is preferably an antigen and this method is useful for
 CC eliciting an immune response from a host against the fusion protein.
 CC The export protein may be a member of the HlyE family such as
 CC cytolysin A (ClyA). The system is useful for the production of
 CC recombinant proteins inside a bacterial host cell, in a bioreactor.
 CC and various devices that facilitate the growth of bacteria. The system
 CC is also useful for providing an animal antigenic material against which
 CC an immune response may be mounted. The obtained recombinant bacterium
 CC (e.g. Salmonella) can be used as a live vector immunogenic composition
 CC capable of facilitating the generation of an immune response
 CC in an animal. The system is also useful for preparing immunogenic
 CC compositions for vaccinating animals, and for use in the food,
 CC nutritional supplement, animal feed, bionediation, waste disposal,
 CC and waste treatment industries. The system is very advantageous for
 CC protein expression, as proteins that are toxic to host bacterium can
 CC also be expressed. A population of recombinant host cells can be
 CC transformed with the expression vector, and then the population of
 CC bacterial host cells can be maintained in culture and used to produce
 CC proteins without having to harvest and lyse the host cells. The present
 CC sequence represents Salmonella typhi ClyA-Bacillus subtilis lethal
 CC levanurase (encoded by sacB gene) fusion protein. The fusion protein
 CC is used to investigate the versatility of ClyA as a fusion partner to
 CC export heterologous antigens out of S. typhi.
 XX SQ Sequence 749 AA;

Query Match 100.0%; Score 1515; DB 24; Length 749;
 Best Local Similarity 100.0%; Pred. No. 1.1e-115;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTSIFAQTVVVKSAIETADGALDLYNKYLQVFPWKTFTDIKLSRFKQYQSEASV 60
 DB 1 MTSIFAQTVVVKSAIETADGALDLYNKYLQVFPWKTFTDIKLSRFKQYQSEASV 60

QY 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVVTOLLISAVILLFDEYNEKKAQKXIDILIRIL 120
 DB 61 LVGDIKVLMDSDQKYFEATQTVYEWCGVVTOLLISAVILLFDEYNEKKAQKXIDILIRIL 120
 QY 121 DDGVKLNKAEQKSLTSSQSFNNASGKLLALDQLTNDTFSEKSSYFQSQVDRIRKEAYAG 180
 DB 121 DDGVKLNKAEQKSLTSSQSFNNASGKLLALDQLTNDTFSEKSSYFQSQVDRIRKEAYAG 180
 QY 181 AAGIVAGPGLIISYSIAAGVIEGKLIPELNRLKTVQNFPTSLSATVQKANKDDAAK 240
 DB 181 AAGIVAGPGLIISYSIAAGVIEGKLIPELNRLKTVQNFPTSLSATVQKANKDDAAK 240
 QY 241 LKLAETIAAIGEIKTETETTRFVYDDLMLSLLKGAAKWINTCNEYQQRHGKKTILFEV 300
 DB 241 LKLAETIAAIGEIKTETETTRFVYDDLMLSLLKGAAKWINTCNEYQQRHGKKTILFEV 300
 QY 301 PDVAS 305
 DB 301 PDVAS 305

RESULT 3

AAR67645

ID AAR67645 standard; Protein; 309 AA.

XX AC

XX AAR67645;

XX AC

XX 25-MAR-2003 (updated)

DT 06-JUL-1995 (first entry)

XX DE

XX Haemolysin regulator.

XX KW

XX Haemolysin regulator; tuberculosis; vaccine; therapy.

XX OS

XX Escherichia coli.

XX PN

XX WO9428137-A1.

XX PD

XX 08-DEC-1994.

XX PF

XX 24-MAY-1994; 94WO-US05869.

XX PR

XX 24-MAY-1993; 93US-0066830.

XX PA

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI

XX King CH, Sathish M, Shinnick TM;

XX DR

XX WPI; 1995-022809/03.

XX DR

XX P-PSDB; AAQ75857.

XX PT

XX DNA encoding carapenem R gene product - used for production of

XX PT carapenem antibiotics in bacteria and for bioassay.

XX FS

XX Claim 9; Page 39-40; 46pp; English.

XX CC

XX The sequence represents the E. coli haemolysin regulator which may

XX CC be used as an immunogen in a vaccine composition against

XX CC tuberculosis (Mycobacterium tuberculosis).

XX CC (Updated on 25-MAR-2003 to correct PN field.)

XX CC (Updated on 25-MAR-2003 to correct PF field.)

XX CC (Updated on 25-MAR-2003 to correct PA field.)

XX SQ

XX Sequence 309 AA;

Query Match 91.7%; Score 1390; DB 16; Length 309;
 Best Local Similarity 91.0%; Pred. No. 6.3e-106;
 Matches 274; Conservative 15; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFAQTVVVKSAIETADGALDLYNKYLQVFPWKTFTDIKLSRFKQYQSEASV 60
 DB 1 MTEIVADKTVVVKNAIETADGALDLYNKYLQVFPWQTFDETIKLSRFKQYQSAASV 60

QY 61 LVGDIKVLMSQDKYFEATQTVYEWGCVVTVLLSAYILLFDEYNEKKASQAQDILIRIL 120
D 120
D 120
QY 121 DGVKKLNEAKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSVQDRIRKAYAG 180
D 180
D 180
QY 181 AAGGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQANKDIDAAK 240
D 240
QY 241 LKLTETAAGGEIKTETTRFYDDMLSLKGAACKQVINTCNEYQORHGKTLFEV 300
D 300
QY 301 P 301
D 301 P 301
RESULT 4
ABG25221
ID ABG25221 standard; Protein; 1035 AA.
AC ABG25221;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #25212.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
FN WO200175067-A2.
XX
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
FR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS89408.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 55580; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits and to assess
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1035 AA;
Query Match 71.2%; Score 1079; DB 22; Length 1035;
Best Local Similarity 92.2%; Pred. No. 9.9e-80;
Matches 212; Conservative 11; Mismatches 7; Indels 0; Gaps 0;
QY 70 MDSQDKYFEATQTVYEWGCVVTVLLSAYILLFDEYNEKKASQAQDILIRILDDGVKKLNE 129
D 129
D 129
QY 130 AOKSILTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQSVQDRIRKAYAGAAAGIVAGP 189
D 189
D 189
QY 190 FGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQANKDIDAAKLTATIAA 249
D 249
D 249
QY 250 IGEIKTETTRFYDDMLSLKGAACKQVINTCNEYQORHGKTLFE 299
D 299
D 299
RESULT 5
ABJ19025
ID ABJ19025 standard; Protein; 2681 AA.
XX
XX AC ABJ19025;
XX
XX 06-MAR-2003 (first entry)
DT
DE Pathogen specific antigen related staphylococcal protein SEQ ID No 274.
KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
KW autoimmune disease; HIV; hepatitis.
XX
OS Staphylococcus sp.
XX
XX WO200259148-A2.
XX
XX 01-AUG-2002.
XX
XX 21-JAN-2002; 2002WO-EP00546.
XX
XX 26-JAN-2001; 2001AT-0000130.
XX
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;
PI Minh DB, Vytvtska O, Etz H, Dryia A, Weichhart T, Hafner M;
PI Tempelmaier B;
XX
XX WPI; 2003-075410/07.
XX
XX Identifying, isolating and producing hyperimmune serum-reactive
PT antigens from a pathogen, for preparing vaccine or medicament for
PT treating or preventing e.g. staphylococcal infections, comprises
PT providing antibody preparation
XX
XX Example 7; Page 200; 252pp; English.
XX
CC The invention relates to a novel method for identifying, isolating and

CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
 CC allergen, a tissue or host prone to auto-immunity, where the antigens
 CC are used in a vaccine, comprises providing antibody preparation from a
 CC plasma pool of a type of animal, or individual sera with antibodies
 CC against the specific pathogen, tumour, allergen, tissue or host prone to
 CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
 CC the 62 sequences of 53-2261 amino acids fully defined in the
 CC specification, or their hyperimmune fragments are useful for the
 CC manufacture of a pharmaceutical preparation, particularly a vaccine
 CC against staphylococcal infections or colonisation against *S. aureus* or *S.*
 CC epidermidis. The preparation of antibodies is useful for the manufacture
 CC of a medicament for treating or preventing staphylococcal infections or
 CC colonisation against *S. aureus* or *S. epidermidis*. The antibody
 CC preparations may also be used for diagnostic and imaging purposes. Other
 CC conditions that can be treated include cancer, autoimmune diseases or
 CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
 CC protozoan pathogens. This sequence represents a staphylococcal protein
 CC relating to the method for identifying and producing pathogen specific
 CC antigens of the invention.

XX
 XX
 SQ Sequence 2681 AA;

Query Match 7.8%; Score 118.5; DB 24; Length 2681;
 Best Local Similarity 21.3%; Pred. No. 2.2;
 Matches 72; Conservative 50; Mismatches 125; Indels 91; Gaps 13;

QY 2 TSIFAEQTVVVKSAIEADGALDLYNKYLQVVPKTFDET-----IKELSR 49
 Db 998 TDLKSESKLDLTKS-YKTLCTSFNDFVYIDEMTMDRIADETEKVNKKYKDYDTLQKNLSD 1056
 QY 50 FQKEY-----SQEASVLVGDIKVL-----LMSQDKYFEATQTVYEWG--G 88
 Db 1057 YMKKYDNCILETSKKYSNDADKVLGDFTAATELQNDQFQVKONWAEFKQTLESFKDG 1116
 QY 89 VTQLLSAYILLDFEYNEKASAKDILIRLDDGVKLINEAKSLTSSQSFNNASGL 148
 Db 1117 IVT-----FAEKARLRVQLDMLDRESMD-----IEEYKSLLANQVNTDIDKNRL 1161
 QY 149 LALDSQLTNDFSEKSSYFOSQVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLI 208
 Db 1162 TA-----SRPFLSVHAS-LRK-----VIEQIADGKVDESEK 1193
 QY 209 PELNRLKTVQNFFTSLSTVQANKDI-----DAAKLKLATEATAAIGEIKTETTRF 262
 Db 1194 TLANNSLNTYNTTLTAYSKTIOEALNTLSQIISSDVASKKVEEFGNVITSSVDVDTIKK 1253
 QY 263 YVD-----YDDLMLSLKGAACKMINTCNEYQQRHGK 294
 Db 1254 QRDGAVITYYSGVPTLSNDPAKSW--TTNDLKDLHIK 1289

RESULT 6
 ABB71280
 ID ABB71280 standard; Protein; 588 AA.

XX
 XX
 AC ABB71280;

XX
 DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 40632.

XX Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

FR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL15383.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 40632; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 588 AA;

Query Match 7.6%; Score 114.5; DB 22; Length 588;
 Best Local Similarity 21.1%; Pred. No. 0.66;
 Matches 67; Conservative 38; Mismatches 97; Indels 115; Gaps 13;

QY 11 EVVKSALIEADGALDLYNKYLQV--IPWKTPEITIKELSRPKQESQASVLVGDIKVL 68

Db 300 EITLALAEANAGALDRFTQLDVSAINKIDAE----- 333

QY 69 LMSQDKYPEA-----TQTVYEW-----CGVTTQLLSAYILLDFEYNEKASQAQ 112

Db 334 -LDNQAEYLDIAEVTBELLRNQTELLMEVELLRGVWTSFQN-----LDIFANRSGIEAV 387

QY 113 KDILIRLDDG-----YKLINEAKSLTSSQSFNNASGLALDSQLTNDFSEKSSY 165

Db 388 SD-LTRLQEQKDRVNLVDKLVNTQGIILRTKGLDD---RLNFVNQLLGVIEPKVNS 443

QY 166 FQSQVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLK----- 216

Db 444 LEDSFDMLNK-----SQINSLEIKNVPEVRLTKTSIRKLSPL 482

QY 217 -----TVQNFFTSLSATV-----KQANK-----DIDAKLKLATEATAAIGEIK 254

Db 483 DNQLALFNQENRYYSVEAVIKAWTPTNLKINDLTHALSISQKRTDLAISGSAEYN 542

QY 255 TETETTRF--YVDYDDL 269

Db 543 TETPTPRFISYKGIEDI 559

RESULT 7
 ABB73774
 ID ABB73774 standard; Protein; 1948 AA.

XX
 AC ABB73774;

XX 30-JAN-2003 (first entry)

XX Candida albicans essential protein SEQ ID NO 7611.

XX Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 XX signal transduction; DNA replication; cell division; growth;
 XX proliferation; Candida albicans; fungicide; antifungal.

XX Candida albicans.

XX WO200253728-A2.
XX 11-JUL-2002.
XX 26-DEC-2001; 2001WO-US49486.
XX 29-DEC-2000; 2000US-259128P.
XX 20-FEB-2001; 2001US-079202A.
XX 22-AUG-2001; 2001US-314050P.
XX (ELIT-) ELITRA PHARM INC.
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX WPI; 2002-566694/60.
XX N-PSDB; ABZ32324.
XX Constructing strains for identifying gene products as effective targets
XX for therapeutic intervention, by inactivating in the strain one allele
XX of a gene and placing other allele of the gene under conditional
XX expression -
XX Claim 44; SEQ ID NO 7611; 167pp + Sequence Listing; English.
XX The invention relates to constructing (M1) a strain of diploid fungal
XX cells in which both alleles of a gene are modified, comprising modifying
XX one allele by insertion or replacement by a cassette having an
XX expressible selectable marker and modifying other allele by
XX recombination, of a promoter replacement fragment with a heterologous
XX promoter, so that expression of the second allele is regulated by the
XX promoter. (M1) is useful for constructing a strain of diploid fungal
XX cells in which both alleles of a gene are modified. The diploid fungal
XX cells having both alleles modified are useful for identifying a gene that
XX is essential to the survival or growth of a fungus, a gene that
XX contributes to the virulence and/or pathogenicity of a fungus, a gene
XX that contributes to the resistance of a diploid fungus to an antifungal
XX agent, an antifungal agent that inhibits the growth of a diploid fungus
XX and for identifying a therapeutic agent for treatment of a mammalian
XX disease. (M1) is useful for identifying a compound which modulates the
XX activity of a gene product, preferably enzymatic activity, carbon
XX compound catabolism, biosynthetic, transporter, transcriptional,
XX translational, signal transduction, DNA replication and cell division
XX activity. The method is useful for identifying a compound having the
XX ability to inhibit growth or proliferation of C. albicans cells and for
XX treating infection by C. albicans. The present sequence is that of an
XX essential Candida albicans protein used in the method of the invention.
XX Note: The sequence data for this patent is not represented in the printed
XX specification but is based on sequence information supplied to Derwent by
XX the European Patent Office.
XX Sequence 1948 AA;
Query Match 7.4%; Score 112.5; DB 23; Length 1948;
Best Local Similarity 22.2%; Pred. No. 4.5;
Matches 58; Conservative 40; Mismatches 90; Indels 73; Gaps 10;
QY 38 KTFDETIKELSRFQOYSCASVLVDIKVILMDSQKYFEATO-----TYVWCVVQ 92
DB 1444 KTLKSTRELBNGSKTEILR-----LKALLRSEDELYQVQENYKTSVHVDYEQDLAQ 1495
QY 93 L-----LSAYILLDFEYNEKASQKDILI--RIILDDGVKKLINEAQS 133
DB 1496 LKVKHETLLSRKNDINESLEYKKSDEYKKLELAESAIAISKEHEQATKEMKESRSQ 1555
QY 134 LITSSQSFNNASGKLLALDSQITWDFSEKSYFQSQVDRIKEAYAGAAAGIVAGPGLI 193
DB 1556 LLLVREE-----LRTTQLIKDFRIKVENLEATIB----- 1566
QY 194 ISYSIAAGVIEGKLIPE-LNNRLKTVQN--FFTSLSATVKQANKDIDAALKLATEIAAI 250
DB 1587 KKHOLDANKEEKIQODKLNHLKVFENKELNKEIKLNRLDLD-----FKTDI--- 1638

QY 251 GEIKTETETTRFFVYVDLML 271
DB 1639 -ETKLIKENKKQLQDVEDVLL 1658
RESULT 8
AAR43563
ID AAR43563 standard; Protein; 476 AA.
XX AC AAR43563;
XX 25-MAR-2003 (updated)
XX 09-JAN-2003 (updated)
XX 05-APR-1994 (first entry)
XX Hyaluronan receptor.
XX Hyaluronan binding protein; HA; RHAMM; mediated motility; wound;
XX healing; diagnosis; treatment; cell locomotion; tumour invasion;
XX birth defects; inflammatory disorder; Alzheimer's disease; dementia;
XX Parkinson's disease; Huntington's disease; AIDS; diabetes; auto;
XX immune diseases; corneal dysplasia; hypertrophy; surgery; burns;
XX strokes; multiple sclerosis; depression; schizophrenia; CNJ;
XX contraception; in vitro fertilisation; embryo development.
XX Rattus sp.
XX WO93211312-A1.
XX 28-OCT-1993.
XX 13-APR-1993; 93WO-CA00158.
XX 09-APR-1992; 92GB-0007949.
XX (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.
XX (UYMA-) UNIV MANITOBA.
XX Turley EA;
XX WPI; 1993-351722/44.
XX N-PSDB; AAQ51212.
XX DNA encoding hyaluronan receptor - used to produce proteins and
XX antibodies for alteration of cell locomotion
XX Claim 7; Fig 23; 88pp; English.
XX The sequence is that encoded by a cDNA clone encoding the hyaluronan
XX receptor (HARC). The sequence was obtd. by screening a 373 library in
XX lambda gtl1 with antibodies to HARC. A clone of 1.9 kb was obtained
XX and used to rescreen the library to obtain the full length, 2.9 kb
XX clone. HA is down regulated in stationary normal cells and is only
XX expressed in situations where cell motility is desired, e.g. in
XX wound healing, in response to growth factors and in chemotaxis by
XX white blood cells. HA may be used for diagnosis and treatment of
XX diseases involving cell locomotion, e.g. tumour invasion, birth
XX defects, acute and chronic inflammatory disorders, Alzheimer's and
XX other forms of dementia, AIDS, diabetes, autoimmune diseases, corneal
XX dysplasias and hypercrotrophies, burns, surgical incisions and adhesions,
XX strokes, multiple sclerosis, depression/schizophrenia related to
XX neuronal growth and pain states involving nerve sprouting; also in CNJ
XX and spinal cord regeneration, contraception, in vitro fertilisation and
XX embryo development.
XX See also AAR46548-51.
XX (Updated on 09-JAN-2003 to add missing OS field.)
XX (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 476 AA;
Query Match 7.3%; Score 110.5; DB 14; Length 476;
Best Local Similarity 20.4%; Pred. No. 1.1;
Matches 68; Conservative 73; Mismatches 125; Indels 67; Gaps 18;

XX SQ Sequence 953 AA;
Query Match 7.3%; Score 110; DB 22; Length 953;
Best Local Similarity 20.6%; Pred. No. 2.9; Indels 90; Gaps 12;
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;
QY 7 EQTEVVVKSATETADGALDLYNKYLDQVWPW-KTFDETIKELSRFKOEY-----SQEA 58
Db 484 EKMEKVKQAKENQETLK--KKLLDQEAQKKELENTALELSQKQFNKAKMLEMAQAN 541
QY 59 SVLVGDIKVLMSODKYFEATQVY--EWCQVVT-----OLLSAYILLFDEYN 105
Db 542 SAGISDAVSRLETNQKQIESLTVHRRRLNDVISIWEKKLNQQAELQBIHQIOEKE 601
QY 106 EKASAKQDILI-----RILDDGVK---LNEAQKSLTSSQSFNNASGKLL 149
Db 602 QEVAELKQILLFGCEKEENKKEITWLKEGVKQDTTLNELQBLKQKSAHVNSLAQDET 661
QY 150 ALDSQL-----TNDPSEKSYFOSQVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVI 203
Db 662 KKAHLEKLEVDLNLKSLKENTFLOELVELKMLAE----- 697
QY 204 EGKLIPELNNRLKTVQNFTSLSATVKQAKNDIDAAL---KLATEIAAIGEI---KTET 257
Db 698 DKRKVSELTSKLTDEEFQSLKSSHEKSNKSLDKSLEPKLSEELAIQDICKCKTE- 756
QY 258 ETRFYVDYDDMLSLKGAAKMIN 283
Db 757 -----ALLEAKTNELIN 768
RESULT 11
ABU53077
ID ABU53077 standard; Protein; 961 AA.
AC ABU53077;
DT 15-APR-2003 (first entry)
DE Intracellular trafficking-associated DKFZphtes3_lg13 homologue #10.
XX Human; gene therapy; vaccine; disease treatment; detection.
XX Homo sapiens.
OS
XX WO2000112659-A2.
XX 22-FEB-2001.
XX 18-AUG-2000; 2000WO-IB01496.
XX 18-AUG-1999; 99US-0149499.
XX 28-SEP-1999; 99US-0156503.
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX Wiemann S;
XX WPI; 2001-327840/34.
XX Nucleic acids having the sequences of clones isolated from libraries of
XX different human tissues, useful in recombinant DNA methodologies -
XX Example III; Page 673; 1095pp; English.
XX This invention describes novel polynucleotides and polypeptides isolated
XX from human cDNA libraries which can be used for gene therapy or in
XX vaccines. The polynucleotides of the invention and antibodies encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. The products of the
XX invention may also be used to identify modulators of expression and
XX activity and to down regulate expression and activity. The antibodies of

CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention.
XX SQ Sequence 961 AA;
Query Match 7.3%; Score 110; DB 22; Length 961;
Best Local Similarity 20.6%; Pred. No. 2.9; Indels 90; Gaps 12;
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;
QY 7 EQTEVVVKSATETADGALDLYNKYLDQVWPW-KTFDETIKELSRFKOEY-----SQEA 58
Db 257 EKMEKVKQAKENQETLK--KKLLDQEAQKKELENTALELSQKQFNKAKMLEMAQAN 354
QY 59 SVLVGDIKVLMSODKYFEATQVY--EWCQVVT-----OLLSAYILLFDEYN 105
Db 355 SAGISDAVSRLETNQKQIESLTVHRRRLNDVISIWEKKLNQQAELQBIHQIOEKE 414
QY 106 EKASAKQDILI-----RILDDGVK---LNEAQKSLTSSQSFNNASGKLL 149
Db 415 QEVAELKQILLFGCEKEENKKEITWLKEGVKQDTTLNELQBLKQKSAHVNSLAQDET 474
QY 150 ALDSQL-----TNDPSEKSYFOSQVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVI 203
Db 475 KKAHLEKLEVDLNLKSLKENTFLOELVELKMLAE----- 510
QY 204 EGKLIPELNNRLKTVQNFTSLSATVKQAKNDIDAAL---KLATEIAAIGEI---KTET 257
Db 511 DKRKVSELTSKLTDEEFQSLKSSHEKSNKSLDKSLEPKLSEELAIQDICKCKTE- 569
QY 258 ETRFYVDYDDMLSLKGAAKMIN 283
Db 570 -----ALLEAKTNELIN 581
RESULT 12
ABU07445
ID ABU07445 standard; Protein; 2230 AA.
AC ABU07445;
DT 28-JAN-2003 (first entry)
DE Protein differentially regulated in prostate cancer #48.
XX Prostate cancer; gene expression; differential regulation;
XX molecular marker; drug target; cancer detection; cancer diagnosis;
XX cancer staging; cancer grading; cancer assessing; cancer monitoring.
XX Homo sapiens.
XX WO200281638-A2.
XX 17-OCT-2002.
XX 08-APR-2002; 2002WO-US10824.
XX 06-APR-2001; 2001US-281731P.
XX 06-APR-2001; 2001US-281732P.
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX Sun Z, Jay G;
XX WPI; 2003-058520/05.
XX N-PSDB; ABX10347.
XX Novel genes which are differentially regulated in prostate cancer,
XX PT useful for diagnosing prostate cancer in prostate tissue sample and
XX PT assessing therapeutic or preventive intervention in prostate cancer
XX patients -
XX Claim 1; Page 299-306; 416pp; English.

XX The invention describes genes (I) which are differentially regulated in
CC prostate cancer. (I) is useful for diagnosing a prostate cancer in a
CC sample comprising prostate tissue, which involves determining the number
CC of target genes which are differentially-regulated in the sample, where
CC the number is indicative of the probability that the sample comprises
CC prostate cancer. (I) is useful for assessing a therapeutic or preventive
CC intervention in a subject having a prostate cancer, which involves
CC determining the expression levels in a sample comprising prostate tissue
CC of target genes which are differentially-regulated in prostate cancer.
CC Preferably, the expression levels of at least 10 genes are determined.
CC (I) is also useful for identifying agents that modulate a biological
CC activity of a polypeptide differentially-regulated in prostate cancer
CC cells, which involves contacting a polypeptide differentially-regulated
CC in prostate cancer cells with a test agent under conditions effective for
CC the test agent to modulate a biological activity of the polypeptide, and
CC determining whether the test agent modulates the biological activity.
CC (I) is useful as molecular markers, as drug targets, and for detecting,
CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC preventing or treating, determining predisposition to diseases and
CC conditions especially relating to prostate cancer. (I) and its expression
CC products are used in the diagnostic test to assay for presence of cancer
CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC blood etc. (I) is useful for assessing cancer e.g., to determine the type
CC of cancer, its stage of development, the nature of genetic defect, etc.
CC The polypeptide encoded by (I) can be used as target for therapy or drug
CC discovery. (I) can also be used for expressing the polypeptide and thus
CC for searching specific binding partners of the polypeptide. (I) is
CC useful in therapeutic applications to treat prostate cancer. The
CC identification of specific genes, and groups of genes, expressed in
CC pathways physiologically relevant to prostate cancer permits the
CC definition of functional and disease pathways and the delineation of
CC targets in these pathways which are useful in diagnostic, therapeutic,
CC and clinical applications. This is the amino acid sequence of a protein
CC differentially regulated in prostate cancer.
XX
SQ Sequence 2230 AA;
Query Match 7.3%; Score 110; DB 24; Length 2230;
Best Local Similarity 20.6%; Pred. No. 8.6; 109; Indels 90; Gaps 12;
Matches 67; Conservative 60; Mismatches 109; Indels 90; Gaps 12;
QY 7 EOTVEVKSATETADGALDLYNKYLDQVIPW-KTFDETIKELSRFKQRY-----SQEA 58
DB 953 EKKMEKVQKAKEMQETLK--KKLLDQBAKLKKELENTALELSQKEQFNAKMLEMAQAN 1010
QY 59 SVLVGDIKVLNDSQDKYFEATQTY--EWCGVVT-----QLLSAYILLFDEYN 105
DB 1011 SAGISDAVSRLETNQKEQIESLTVHRRELNDVSIWEKKLNQQAELQEIHLQLOKE 1070
QY 106 EKKAQAQDKILT-----RIILDGIVKK--LNEAQKSLTSSQSFNNASGKLL 149
DB 1071 QEVAELKQILLFGCEKEEMKEITWLKEGVKQDTTLNELQELKQKSAHVNSLAQDET 1130
QY 150 ALDSQL-----TNDFSEKSSYFQSDVRIRKEAVAGAAAGIVAGPGLIISYIAQVI 203
DB 1131 KKAHLEKLEVDLNLKSLKENTLQELQELVLMKLAEE----- 1166
QY 204 EGKLIPELNNRLKTVQNFTSLSATVKQAKDIDAQKL---KLATEIAAIGE---XTET 257
DB 1167 DKRYSELTSLKLTDEEPQSLKSSHEKSLDKSLSEFKLSLELAQLDICKKTE- 1225
QY 258 ETRFYVDYDDMLSLKGAAKMIN 283
DB 1226 -----ALLEAKTNELIN 1237
RESULT 13
AAB96493
ID AAB96493 standard; Protein; 739 AA.
XX
AC AAB96493;
XX

DT 29-OCT-2001 (first entry)
XX Putative sensory transduction histidine kinase and response regulator #3.
DE
XX Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
EN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-0005034.
XX
PR 21-APR-1999; 99FR-0005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST PR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
DR WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
XX
PS Claim 7; Pages 1203-1205; 1657pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAB96431 and AAB91223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO200065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAB75903-AAB75920 and AAB66436.
XX
SQ Sequence 739 AA;
Query Match 7.2%; Score 109; DB 22; Length 739;
Best Local Similarity 19.2%; Pred. No. 2.5;
Matches 64; Conservative 71; Mismatches 138; Indels 50; Gaps 11;
QY 1 MTSIFAEQVEVVKSA--JETADGALDLYNKYLDQVIPWKTDFETIKELSRFKQYSQEA 58
DB 425 VTETFRSIGSLVEMANDLEKRNALAQVSKDVTET--NQVNEATQVYSIEAQRQETI 482
QY 59 SVLVGDIKVLNDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAKAQDKI--- 115
DB 483 NEITDGMRLVAQTS-----ESVRAKEEFGSGAVTEVVS-----IANEGSQKDEALKKIEDI 534
QY 116 --LIRILDGVKLNIAQKSL-----LTSOSFNNASGKLLALDS----- 153
DB 535 QHVMRSIETVSKVAEMSRNIBETITNVITSIAEQTN-----LLALNAAIEARAGEAGRGF 590
QY 154 -----QLTNDPSEKSSYFQSDVRIR---KEAVAGAAAGI-VAGPGLIISYIAAG 201
DB 591 AVVAQIRKLAESKQAADNIKSIIDKITDEIKEAVEATKEGVSVTGESSEITLDTIGYL 650
QY 202 VIEGKLIPELNNRLKTVQNFTSLSATVKQAKNDIDAQKLKLAETIAAIGEIKTETETTR 261
DB 651 ANIATLLQETSERMTTKE-----QIVRTQBEVDKALLENLAASAETTAGAEVVS 703
QY 262 FVYDYYDDMLSLKGAAKKMINTCNEYQQRHGK 294
DB 704 SAIEQQTAAIEELRRAAQELKDMVGRMROIVGK 736
RESULT 14
AAB96493

ABU53079 standard; Protein; 929 AA.
ABU53079;
15-APR-2003 (first entry)
Intracellular trafficking-associated PKF2phtes3_1g13 homologue #12.
Human; gene therapy; vaccine; disease treatment; detection.
Homo sapiens.
WO200112659-A2.
22-FEB-2001.
18-AUG-2000; 2000WO-IB01496.
18-AUG-1999; 99US-0149499.
28-SEP-1999; 99US-0156503.
(GEHU-) GERMAN HUMAN GENOME PROJECT.
Wiemann S;
WPI; 2001-327840/34.
Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies -
Example III; Page 674-675; 1095pp; English.
This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention.

Query Match 7.2%; Score 109; DB 22; Length 929;
Best Local Similarity 20.6%; Pred. No. 3.4;
Matches 67; Conservative 59; Mismatches 106; Indels 94; Gaps 12;
Qy 7 EQVEVVKSAIETADGALDLYNKYLDQVWP-KTFDETIKELSRFKQEV-----SQEA 58
Db 1 EKVKQAKEMQET-----LKKLLDOEAKLKELENTALELSQKEKQFNKMLEMAQAN 54
Qy 59 SVLVGDIKVLIMDSQDKVFATQTVY--EMGVVVT-----QLLSAYILLFDEYN 105
Db 55 SAGISDAVSRLETNQKEISLEFVHRRLNDVSIWEKLNQOAELOEIHQIQEKE 114
Qy 106 EKKAQAKDILI-----RILDDGVK---LNEAQSLTSSQSFFNNSGKLL 149
Db 115 QEVAELKQKILLFGCEKEEMNKETTLKKEGVQDITLNEQLQKQSAHVNSLAQDET 174
Qy 150 ALDSOL-----TNDPSEKSYQSQVDRIRKEAYAGAAAGVAGPGLIISYIAAGVI 203
Db 175 KLKALEKLEVDLNSKENIFLOEQLVELKMLAE----- 210
Qy 204 EGKILPELNRLKTVQNFVTSATVKQAKNDIAAKL-----KLATEIAIGEI---KTET 257
Db 211 DKRVSELTSLKLTDEFOSLGSKSEKSKSLDKSLFKLSELAIQDICKCKTE- 269
Qy 258 ETTTFYVDYDDLMLSLKGAQKMIN 283
Db 270 -----ALLEAKTNELIN 281

RESULT 15
ABB59245
ID ABB59245 standard; Protein; 1048 AA.
XX AC ABB59245;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 4527.
XX DE Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li FWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL03348.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX PT Disclosure; SEQ ID NO 4527; 2ipp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AB57737-AB572072).
XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Query Match 7.2%; Score 108.5; DB 22; Length 1048;
Best Local Similarity 21.6%; Pred. No. 4.3;
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;
Qy 24 LDLINKYLDQVIPKMTFDETIKELSRFKQESQASVLVDIKVLLMDSQDKYFEATQV 83
Db 159 LELYNELCDLL-----STDDITK--IRIFDSTKSGSVIOGLEIPVHSKDDVYKLEKG 213
Qy 84 YEWCGVVTQLLSAYILLFDEYNEKKAQAQKILIRILDDGVK-----KLNEAQKSLTSS 138
Db 214 KERRKTATLNA-----QSSRSHVTSIVVHIRENGIEGEDMLKIGKLNVLVLAGS 265
Qy 139 QSFNNASK-----LLALDSQLTN--DFSEKSSYQSQVDRIRKEAYAGAA 182
Db 266 ENVKAGNEKGIKRVETVNIQSLLTGRVITLVDRAPHVPVRESKLTLLQESLGRT 325
Qy 183 -----AGIVAGPFGI-----IISYSIAAGVIEGKLIPELNNRL--KIVQNFFTSLSATV 230
Db 326 KTSIIATISPGHKEIEETLSTLEYAHRKQIQNK--FEVQKLTKTIVLKEYTE--EID 380
Qy 231 QANKDIDAAKUKLATEIA--AIGEIKTETETTRFYVDYDDLMLSLK 275

Db 266 ENVSKAGNEKGIKRVETVAINQSLTLGRVITLVDRAPHPVRESKLTRELSLGRT 325
Qy 183 -----AGIVAGPGL-----IISYSTAAGVIEGKLPELNNRL--KTQNPFTSLSATVK 230
Db 326 KTSIIATISPGHKDIBETLSTLYAHRAKVIQNK--FEVNOKLTKTKVLKYTE---EID 380
Qy 231 QANKDIDAAKLKATEIA--AIGIKTETETTRFYVDYDDMLSLK 275
Db 381 KLRDLNAROKGKIYLAETTYGEITLKLESQRELNEKVLKALK 427

RESULT 18

ABP40235
ID ABP40235 standard; Protein; 3696 AA.

XX AC ABP40235;

XX DT 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5080.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX PD 30-APR-2002.

XX PF 13-AUG-1998; 98US-0134001.

XX PR 14-AUG-1997; 97US-055779P.

XX PR 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX PI Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

XX N-P5DB; ABN92780.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX disclosure; SEQ ID 5080; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.

XX SQ Sequence 3696 AA;

Query Match 7.2%; Score 108.5; DB 23; Length 3696;
Best Local Similarity 21.5%; Fred. No. 22;
Matches 71; Conservative 49; Mismatches 128; Indels 83; Gaps 12;

Qy 11 EVVKSATETADGALDLYNKLDVLPWKTFDTIKELSR-----FKQYISQ-EA 58
Db 568 EQVNDIIPSNVTLASYNKY-----NKLKERAQTVLDBETNNTPNQYSQTQI 616

Qy 59 SULVGDIKVLIM-----DSQDKFEATQIVYEWGCVVTVLLSYILLPEYNEKASA 111
Db 617 DLLLHELQTLINRVASREINDKAQMTDAVIDSTELTTEKDT---LVDOENHKNKI 673

Qy 112 QKDLIRILDGVKKLNEAOKSLITSS-----QSFNNASGKLLALDSQLTNDPS 160
Db 674 SNIDDELTDGVERVKEAGLHTLES DTPHPTKPNARQVNNRA-----DQOKTILRN 727
Qy 161 EKSSYFOSQVDRIK-PAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELANNRLKTQV 219
Db 728 NHEATTEQNEAIFQVEAHSDAIKIGE-----AETDTTVNEAR 767
Qy 220 NFFTSLSGAT-VKQANKDIDAAKLKLATEIAAIGIKTETETTRFYVDYDDMLSLKGA 278
Db 768 DNGTKLIATDVPNPTKAEK---RAAVTNSANSKIKDINNNTQATLDERNDALAVNRSK 824
Qy 279 KKMINTONEYQ-----QSHGKKTLPFVP 301
Db 825 DEAIQNTAQGNDVDVTEAQNNGTNTIOQVP 855

RESULT 19

ABB77986
ID ABB77986 standard; Protein; 1292 AA.

XX AC ABB77986;

XX DT 22-OCT-2002 (first entry)

XX Amino acid sequence of an Arabidopsis RAD50 homologue.

XX Nucleic acid integration; homologous recombination; telomeric region;
XX RAD50.

XX Arabidopsis thaliana.

XX EP1217074-A1.

XX PD 26-JUN-2002.

XX PF 22-DEC-2000; 2000EP-0204693.

XX PR 22-DEC-2000; 2000EP-0204693.

XX (UTLE-) RIJKSUNIV LEIDEN.

XX (BINA-) STICHTING BINAIR VECTOR SYSTEEM.

XX Hooikaas PJJ, Van Attikum H, Bundock P;

XX WPI; 2002-550409/59.

XX Directing integration of nucleic acid of interest to a sub-telomeric
XX region in an eukaryote with preference for non-homologous
XX recombination, by steering an integration pathway towards homologous
XX recombination -

XX Disclosure; Fig 5; 63pp; English.

XX The specification describes a method for directing integration of
XX a nucleic acid of interest to a pre-determined site, where the nucleic
XX acid has homology at or around the pre-determined site, in a eukaryote
XX with a preference for non-homologous recombination. The method comprises
XX steering an integration pathway towards homologous recombination. The
XX method is useful for directing integration of a nucleic acid of interest
XX to a subtelomeric and/or telomeric region in an eukaryote with a
XX preference for non-homologous recombination. The nucleic acid of
XX interest comprises an inactive gene to replace an active gene, or vice
XX versa, is a portion of a gene delivery vehicle, confers a desired
XX property to the eukaryotic cell, or encodes a therapeutic proteinaceous
XX substance or a substance conferring resistance for an antibiotic
XX substance to a cell. The method is useful for improving gene targeting
XX efficiency. The method is useful in the replacement of
XX an active gene by an inactive gene, for e.g. for the inactivation of
XX genes controlling undesired side branches of metabolic pathways, to
XX increase the quality of bulk products such as starch, or to increase
XX the production of specific secondary metabolites or to inhibit

CC formation of unwanted metabolites, and also to inactive genes
 CC controlling senescence in fruits and flowers or that determine flower
 CC pigments. The method is also useful for replacing an inactive gene by
 CC an active gene. For e.g. the replacement of a defective p53 by an
 CC intact p53. Many tumours acquire a mutation in p53 during their
 CC development which renders it inactive and often correlates with a poor
 CC response to cancer therapy. By replacing the defect p53 by an intact
 CC p53, e.g. through gene therapy, conventional anti cancer therapy have
 CC better changes of succeeding. The method is also useful for therapeutic
 CC prodrugs substance integration. A tumoricidal gene can be
 CC delivered to a pre-determined site present only in e.g. proliferating
 CC cells, or present only in tumour cells, e.g. to the site where a tumour
 CC antigen is expressed form. ABB77984-86 represent RAD50 homologues. RAD50
 CC is involved in non-homologous recombination.

XX Sequence 1292 AA;

Query Match 7.1%; Score 107.5; DB 23; Length 1292;
 Best Local Similarity 17.2%; Pred. No. 6.8;
 Matches 69; Conservative 77; Mismatches 138; Indels 117; Gaps 13;
 QY 2 TSIPAEQTVVVKSAIETADGALDLYNKYLQVVPWKT-FDE----- 42
 Db 281 STLFKEQQRQVAALPEEED-----TIELKWKSKFEERLALLOTKIRKQWERMD 332
 QY 43 ---TINKLSRFKQYSQASVL--VGDIKVLMDSQDKYFEATQTVYEWCGVVTQLLSAY 97
 Db 333 TETTISLHNKNTNMLEISKLOTEAEAHMLLKNERDSTIQNIFFHYNLGNVPSTPFSTE 392
 QY 98 ILLPFEYNKKAQAKDILIRLDGVKKLNEACKSLTSSQSFNNASGKLLALDSQ--- 154
 Db 393 VVL--NLNTRIKSRIGELEMILLDK--KKSNET--ALSTAWDCYMDANDRWKSEIAQKRA 446
 QY 155 -----LTNDFSEKSSY-----FQSQ 169
 Db 447 KDEIKWGISKRIEKEIERDSFEPISTVDVKQTDREKQVQVELERKTKQNSERGESEK 506
 QY 170 VDIRKEAYA-----GAAGIVAGPFGLIISYI-----RAGVIEGKLIIFELNLR- 214
 Db 507 IEQOHEIYSLBHKITLNRERDVNAGDAERLRLTRIDECKDRINGVLKGLRPEKQMKR 566
 QY 215 ----LKVQNFPTSLSATVQKANDIDAALKLATEIAAIGETETETETTRFYVDYD-- 267
 Db 567 EIVQALRSIEREYDDLKLSREAEKVNLMQKIOEVNNSLFKINKDTSRKRTIESKQ 626
 QY 268 -----DMLSLKGAQKQKQNTCNEYQORHGKTLFE 299
 Db 627 ALKQESVTIDAYPKLLESKAKDRDRKREYNMANGRMQMF 667

RESULT 20

ABP39975
 ID ABP39975 standard; Protein; 1211 AA.
 AC ABP39975;

24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4820.

KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy.

OS Staphylococcus epidermidis.

PN US6380370-B1.

PD 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

DR N-PSDB; ABN92520.

PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis

PT polypeptide, useful for diagnosing and treating bacterial infections -

PS Disclosure; SEQ ID 4820; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.

CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

XX Sequence 1211 AA;

Query Match 7.1%; Score 107; DB 23; Length 1211;
 Best Local Similarity 18.3%; Pred. No. 6.9;
 Matches 61; Conservative 59; Mismatches 131; Indels 82; Gaps 9;

QY 7 EQTVVVK-----SAIETADGALDLYNKYLQVVPWKTFTETIKELSRFKQYSQASV 60
 Db 185 EESAGVLTKYKKAESIQKLDHTEDNLNRVEDILYDLGRVPLKEEAIAKEAYQLSKSE 244

QY 61 LVGDIKVLAMDSDQKYFEATQTVYWC-----GVVTQLLSAYILLPDEYNE 106

Db 245 MEQSDVIVTVSDIDHTEDNQRLDERNLHLSQQAQKEQQAQINQLQKY----- 295

QY 107 KKAQAKDILIRLDGVKKLNEACKSLTSSQSFNNASGKLLALDSQTNP-----FS 160

Db 296 -KGKQON-----DYDIEKLN--YELVKATENYEQLSGKLNLEERKKNQSETNARYE 345

QY 161 EKSSYFQSQVDIRKEAYAGAAAGIVAGPFGLIISYIAAGVIEGKLIPELNNRL----K 216

Db 346 EELDNLSEQIDSINKKAQNE-----KLJADLNKQKQNLNK 381

QY 217 TVQNFPTSLSATVQKANDIDAALKLATEIAAIGETETETETTRFYVDYDMLSLKSG 276

Db 382 EVQELSSLVYSDEQHQDEKLEIKNSYYTLMSEQSVVNDIRFLEHTINENAKKSRDLS 441

QY 277 AAKKQNTCNEYQO-----RKGKTLFEV 300

Db 442 RLVEAFNQLKDQQNITQTKQYQSSKSKMEKV 474

RESULT 21

ABB49720

ID ABB49720 standard; Protein; 927 AA.

AC ABB49720;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #2424.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KW vitamin B12; bacterial infection; disease.

OS Listeria monocytogenes.

XX WO2001/7335-A2.

XX

PD 18-OCT-2001.
XX 11-APR-2001; 2001WO-FR01118.
XX 11-APR-2000; 2000FR-0004629.
XX (INSP) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Feihi H, Dehoux P;
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Coscart P;
PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain I, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
PT and prevention of *Listeria* and related bacterial infections, and
PT related polypeptides -
XX
XX Claim 6; SEQ ID No 2425; 192pp; French.
XX The present invention relates to the genome sequence of *Listeria*
CC *monocytogenes* EGD-e (see ABA3041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in *L.*
CC *monocytogenes* and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of *L. monocytogenes* and related organisms,
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by *L.*
CC *monocytogenes* and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 927 AA;
Query Match 7.0%; Score 106.5; DB 23; Length 927;
Best Local Similarity 20.5%; Pred. No. 5.4;
Matches 78; Conservative 44; Mismatches 131; Indels 127; Gaps 16;
QY 1 MTSIFAQTVVVKSAIETADG---ALDLNKNYLDQVWPV---KTFDETIKELSRFQOE 53
DB 160 VTKSYAEAFDKIK---ESGDGFAQAADGGKINDGLVKSQEGNKITSTNLKTLADSSLT 216
QY 54 YSQEASVLI-----VGDIK-----VLLMDS 72
DB 217 FKDGANTLEVGLKTYTDGNTAAAGCKLNAGVSTLAAGVPLKDGVAALDGGATKLASG 276
QY 73 QDKYFEATQTVYVCGVVTQLLSAYILLFDEYNEKKASAKQILIRILDGVKKXNEAQK 132
DB 277 VSTYTSQVDTL---AGGINQAYTGSTALSDGLNKNWS-----VPTLASSGITQLNNGQK 327
QY 133 SLTSSOSFNNAKGLLA---LDSQLTN-----DFSEKSSYFOSOVDRIRAEVAGRA 192
DB 328 SLATGLSLVDGSKNLKAGLKELDGNJTDQKQIAQLKQGNLDLQOQIDQLNOSVNGEDA 397
QY 183 AGIVAGPGLIISYIAAGVIEGKLIPELNNRL-----KTVQNFTLSAT 228
DB 388 A-----LAKQAT---LQKSLDQLNGLTFIKSNANFAEAIKSKINATAGVSAE 434
QY 229 VKQ-----ANKDIDAALKLATEITAIGEIKT-----ETETTRFYVDVDDLMSL 273
DB 435 DRKIIDAIQADLDKETQKATQVATVEQJGSLGDLAAIQVNT-----EL 483

QY 274 LKGAACKMINTCNEYQQRHG 293
DB 484 QTGVAKISAG-----YQAVHG 499
RESULT 22
AAW22775
ID AAW22775 standard; Protein; 1312 AA.
AC AAW22775;
XX 21-DEC-1998 (first entry)
XX Human RAD50.
XX Human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;
XX central nervous system.
XX Homo sapiens.
XX WO9727284-A2.
XX PD 31-JUL-1997.
XX 24-JAN-1997; 97WO-US01299.
XX 17-JUL-1996; 96US-0687080.
XX 26-JAN-1996; 96US-0592126.
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX Dolganov G;
XX WPI; 1997-393672/36.
XX N-PSDB; AAT75237.
XX Human tumour suppressor gene RAD50 - useful to detect
PT predisposition to, decrease risk of and treat cancer, also Septin-2
PT homologues
XX Claim 5; Page 82-86; 195pp; English.
XX The human RAD50 (hrRAD50) is involved in DNA repair and has tumour
CC suppression activity, can be used to detect predisposition to, decrease
CC the risk of or treat cancers, e.g. acute myeloid leukaemia,
CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,
CC therapy related acute myeloid leukaemia, refractory anaemia or refractory
CC anaemia with excess blasts. Also disclosed in this invention is Human
CC Septin-2 homologues of which may be used as targets for cancer therapies
CC and central nervous system directed treatment methods, and to measure the
CC proliferative potential of selected cell types.
XX
SQ Sequence 1312 AA;
Query Match 7.0%; Score 106.5; DB 18; Length 1312;
Best Local Similarity 18.9%; Pred. No. 8.4;
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;
QY 10 VEVKSAIETADGALDLYNKYLDQVWPVKTDETIK-----ELSRFQOEY 54
DB 502 MEVISLQNEKAD--LDRTLRLKLDQEMEQLNHHHTTQMEMLTKDKADQEKIRKISRH 559
QY 55 SQEASVIVGDIKVLMSQSKYFEATQTVYVCGVVTQLLSAYILLFDEYNEKKASAKQD 114
DB 560 SDELTSLLG-----YFNKKQLEDWLSKSK-----EINQTR----- 591
QY 115 ILIRILDGVKKXNEAQKSLTSSQFNNAKGLLALDLSQTLNDPSEK-----SSYFQS 168
DB 592 -----DRLAKLN---KELASSEQNKHNNELKKEEQLSS-YEDKLPVCGSQDFES 640
QY 169 QVDRIRKE-----AYAGAAGIVAGPF----- 190

Db 641 DLRLKEIEKSKQKRAMLAGATA--VYSQFITQLTDENQSCPCVQRFQTEAELOEVI 698
 QY 191 -----GLI-ISYSIAAGVIEGKLIPELNNRLKT 217
 Db 699 SLLQSKLRAPDKLKSTESSELKKRRDEMGLVPMRQSIID--LKEKEIPELRNKLQN 756
 QY 218 VQNFFTSLSATYQKANKIDAAKIKLATEIAAIGETETETTT-----RFYVDY 266
 Db 757 V-----NRDIQRLKNDIEEQETELGTIMPEESAKVCLTDVTIMERFOMEL 802
 QY 267 DDLMLSLKGAAK-----KMINTCNEYQQRHGK 295
 Db 803 KQVERKIAQQAALQIGIDLDRITVQVNOBKQKQHK 838

RESULT 23
 AAW71295
 ID AAW71295 standard; Protein; 1312 AA.
 XX AAW71295;
 XX 25-NOV-1998 (first entry)
 XX Human homologue of yeast RAD50.
 XX Human homologue; yeast RAD50; Drosophila Septin-2; Acyl-CoA synthetase;
 KW immunomodulatory activity; identification; activated T-cell.
 XX Homo sapiens.
 XX WO9838306-A1.
 XX 03-SEP-1998.
 XX 27-FEB-1997; 97WO-US03159.
 XX 27-FEB-1997; 97WO-US03159.
 XX (GENE-) GENELABS TECHNOLOGIES INC.
 XX Dolganov G;
 PI WPI; 1998-481207/41.
 DR N-PSDB; AAV59979.
 PT Novel human immunomodulatory polypeptide(s) - have homology to the
 PT yeast RAD50 or Drosophila Septin-2 proteins
 XX Disclosure; Pages 136-140; 155pp; English.
 XX The present sequence represents a human homologue of the yeast
 CC S. cerevisiae gene RAD50. The present sequence has 35% overall
 CC homology to the yeast RAD50 gene, and is expressed in activated
 CC T-cells, testis, foetal liver and heart tissues. The specification
 CC also describes sequences encoding human homologues of the
 CC yeast RAD50, the Drosophila Septin-2 and Acyl-CoA synthetase. The
 CC proteins have immunomodulatory activity. The nucleic acids and
 CC population can be used to identify activated T-cells in a sample
 CC encoding other proteins or other compounds having immunomodulatory
 CC activity.

Query Match 7.0%; Score 106.5; DB 19; Length 1312;
 Best Local Similarity 18.9%; Pred. No. 8.4;
 Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;
 QY 10 VSWKSAITAGALDLYNKYLDOVIPKTFDETETIK-----ELSRFKQY 54
 Db 502 MEVISLQNEKAD--LDRTLRKLDQNEQNLNHTTTRTQMEMLTKADKDEQIRKISRH 559
 QY 55 SOEASVYLVGDIKVLNMDSDQKFEATQTVYEWGVVTVLLSAYILLFDEYNEKASAKQD 114

Db 560 SDELTSILG-----YFPNKKQLEDMLHKS 591
 QY 115 ILIRIILDDGVKLNENAKSLTSSQSFNNASGKLALDSQLTNDPSEK-----SSYFOS 168
 Db 592 -----DLAKLN--KELASSEQNKHNINNELKREBQLSS-YEDKLFVCGSQDFES 640
 QY 169 QVDRIRKE-----AVAGAAAGIVAGPF----- 190
 Db 641 DLRLKEIEKSKQKRAMLAGATA--VYSQFITQLTDENQSCPCVQRFQTEAELOEVI 698
 QY 191 -----GLI-ISYSIAAGVIEGKLIPELNNRLKT 217
 Db 699 SLLQSKLRAPDKLKSTESSELKKRRDEMGLVPMRQSIID--LKEKEIPELRNKLQN 756
 QY 218 VQNFFTSLSATYQKANKIDAAKIKLATEIAAIGETETETTT-----RFYVDY 266
 Db 757 V-----NRDIQRLKNDIEEQETELGTIMPEESAKVCLTDVTIMERFOMEL 802
 QY 267 DDLMLSLKGAAK-----KMINTCNEYQQRHGK 295
 Db 803 KQVERKIAQQAALQIGIDLDRITVQVNOBKQKQHK 838

RESULT 24
 ABB63519
 ID ABB63519 standard; Protein; 2346 AA.
 XX ABB63519;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 17349.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL07622.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 17349; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

```
SQ Sequence 2346 AA;
Query Match 7.0%; Score 106.5; DB 22; Length 2346;
Best Local Similarity 18.9%; Pred. No. 18;
Matches 58; Conservative 66; Mismatches 118; Indels 65; Gaps 11;

QY 25 DLYNKYLDQVPMKTFETIKELS---RPFQYSQBSASVLVGDIKVLLMDSQDKYFEATQ 81
DB 203 ELQNRREHTINTWLOSLCKEXTESLKMGEQYQAVKTIGELTISKIEMONDFAKQNO 262
QY 82 TVVWCGVTVTLLSAYILLDFEYNEKKAS--AQKDILRLDDGVKKLNEAK--SLITS 137
DB 263 ATBEYVGLKXELDAKELPEIFAKSTESDHLIQRELLQGISIKELLEAEQCAQLTE 322
QY 138 -----SQSFNNASGKLLALDSQL--TNDF--SEKSSYFQSQVDRIKRAYAGAAAG 184
DB 323 QMETMKQHSALDEQNKIQAMEQBLASANDLLKQARESNIESAICQLAPSA----- 375
QY 185 IVAGPGLIISYSIAGVIEGKLIPELNNELKTQVFFTSLSATVQKANKDIDAAGKLA 244
DB 376 -----AVAGSLRSDI-----SLTELYSMYAKSSEELMRNCEIEQKLOLQ 417
QY 245 TEIAAIGEI-----KTETETRFYVDYDDIMLS--LLKGAAKKMINTCNE 287
DB 418 SIIAIESAPILEKQNSDYQKMKETNSLLR---EHDLLQNKLCLELERALSTLNH 474
QY 288 YQQRHGK 294
DB 475 NQNEKK 481

RESULT 26
ABB71136
ID ABB71136 standard; Protein; 7201 AA.
XX
AC ABB71136;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 40200.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PS CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL15239.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 40200; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
```

```
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB957737-AB972072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7201 AA;
Query Match 7.0%; Score 106.5; DB 22; Length 7201;
Best Local Similarity 18.9%; Pred. No. 77;
Matches 68; Conservative 51; Mismatches 131; Indels 109; Gaps 14;

QY 18 ETADGA-LDLYNKYLDQVIEP-----KTDETIKEL 47
DB 3504 QEAQGVQLSSYQDILNQTAVNLDQVEKLIHNENPASWTSQAERSKLYKYKATNQDINSH 3563
QY 48 SRFQEQYSQASVLVGDIKVLLMDSQDKYFEATQTVYEWCG-----VVTOLL SAYILL-- 100
DB 3564 KRIVEAVNEKKAALLGSAAPANADEISKAVAENVKRYDQVQDCAKLVADLDGAFVYQ 3623
QY 101 FDEYNEKKAQAQKIDILIRI-----LDGKVKLNEAKSLTSSQSFNNASGKL 148
DB 3624 FSELQKACQDYQKNLMDRLTGYSYSGNKAALQARLQKINEIQAL-----PEGVAKL 3676
QY 149 LALDSQLTNDFSEKSSYFQSQVDRIKRAYAGAAAGIVAG--PGLIISYSIAGVIEGK 206
DB 3677 KSLSDHIEQQ-----ASNIPARSEVWARDLIANLHADPEKFGASLS-DVKSG-LENR 3726
QY 207 L-----IPELNNRLKTQ-----NFTTSLSATVQKANKDIDA 238
DB 3727 LQWWDYEINLDRILTWLGEAENSLKNYLNKSSPEEKEEQLNGFQSLAQLRQNEADFDK 3786
QY 239 AKLKLATEIAAIGE-----IKTETETTFYVDYDDLLSLKGAAKKMINTCNEVQQRHG 293
DB 3787 VKDTSSELVSSGETRIAVNVQVSSRP-----QSIQATAKEILKKEQAVQDHHG 3836

RESULT 26
ABB64018
ID ABB64018 standard; Protein; 685 AA.
XX
AC ABB64018;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 18846.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PS CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL08121.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
```


CC proteinaceous substance integration. A tumoricidal gene can be
CC delivered to a pre-determined site present only in e.g. proliferating
CC cells, or present only in tumour cells, e.g. to the site where a tumour
CC antigen is expressed from. AB377984-86 represent RAD50 homologues. RAD50
CC is involved in non-homologous recombination.

XX Sequence 1318 AA;

Query Match 7.0%; Score 105.5; DB 23; Length 1318;
Best Local Similarity 18.4%; Pred. No. 10;
Matches 72; Conservative 53; Mismatches 106; Indels 161; Gaps 14;
QY 10 VEVVKSATETADGALDLYNKYLQVLPWKTFDETIK-----ELSRPKQEF 54
Db MEVISLQNEKAD--LDRTLRKLDQMEQLNHHHTTRTQMEMLTKDKADKDEQIRKIKSRH 565
QY 55 SQEASVLVDGDKVLMDSDQKYFATQVYEWCGVVTQLLSAYILLFDEYNEKKASAKD 114
Db SDELTSLG-----YFPNKKQLEDMLHKSCK-----EINQTR----- 597
QY 115 ILIRILDGVKLNKAEAKSLTSSQSFNNASGKLLALDSQLTNDFSEK-----SSYFQS 168
Db -----DRLAKLN--KELASSEQNKHNINNELEKEEQUSS-YEDKLFVCGSQDPFES 646
QY 169 QVDRIRKE-----AYAGAAA-----GIVAGPFGULIISYSIAAGV 202
Db DLDRLKEIEKSKQORAMLAGATVYSOFITQLTDENQSCPCVQCFQTEAEIQEALSD 706
QY 203 IEGKL-----IPENNRKLTQVNF 221
Db LQSKRLAPDKLKSTESLKKKRRDEMLGLAPMQRQSIIDLKEKEIPELENKLNQV--- 763
QY 222 FTLSATVKQAKOIDAAKLATEIAAIGIKETETTT-----RFYVDYDLM 270
Db -----NRDIQLRKNIDIEQETLLGTIMPEESAKVCLTDVIMERFQWELKDVE 812
QY 271 LSLKGAAG-----XMINTCNYQQRHGKK 295
Db 813 RKIAQAAKLOGIDLDRTVQVQVQNEKQKQKH 844

RESULT 30

ABP76825
ID ABP76825 standard; Protein; 1015 AA.

XX AC ABP76825;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae amino acid sequence SEQ ID 180.
XX KW Antibacterial; infection; vaccine; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB02069.
XX PR 12-FEB-2001; 2001GB-0003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Massignani V, Monaci E;
XX DR WPI; 2003-058415/05.
XX DR N-PSDB; AB377795.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
XX

PS Claim 1; Page 197; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.

XX Sequence 1015 AA;

Query Match 6.9%; Score 105; DB 24; Length 1015;
Best Local Similarity 21.5%; Pred. No. 8;
Matches 74; Conservative 55; Mismatches 113; Indels 102; Gaps 17;
QY 9 TVEVKSATETADGALDLYNKYLQVLPWKTFDETIKELSRFKQESQEASVLVG 63
Db 104 TADFAKSLVDLTDAD-AMQSINSQVRQVTSSETYEYLAQQQLDNTANTRASLESTANLYVS 162
QY 64 DIKVL--MDSQD--KYFEATQTYVWCGV-----VTQLLSAYILLFDEYNEK 107
Db 163 TGRALKDYGTQOEILKFTFEEAANNAMTIGVGAGQQAALMQLSQALGSGVLQGEF--K 220
QY 108 KASAKDILIRL-----DGVKYL-NEAQ-----KSLTSSQSFNNASGKLLALD 152
Db 221 SISEAAPILLDTIAEWMGKSRDEIKLGSEGLTADVIFKAIKSGASEKFGQAAKMPVTM 280
QY 153 SOLTNDSEKSSYFQSQVDRIRKE-----AYAGAAA- 183
Db 281 GQALTVFSNN---WQSMVSKLMDSGTMSGIAAVIKLIADNLNVVPIVAGFAVAVA 337
QY 184 -----GIVAGPFGULIISYSIAAGVIEGKLIPELNNRLKTVQNFSTLSATVKQAKND 235
Db 338 APTLAINLALLANPFGII---AVAGTVVG-LIAKFGDEIDVFGGWSNLS-----D 385
QY 236 IDAAKCLKATEIA--AIGEIKTETETTRFYVDYDLMLSLKGA 277
Db 386 VIRAVWQLITETVGEAVGVTKSW-----FDGLTGRNLNEGA 420

RESULT 31

ABP78617
ID ABP78617 standard; Protein; 1015 AA.

XX AC ABP78617;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae amino acid sequence SEQ ID 3764.
XX KW Antibacterial; infection; vaccine; gene therapy.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB02069.
XX PR 12-FEB-2001; 2001GB-0003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Massignani V, Monaci E;
XX DR WPI; 2003-058415/05.
XX DR N-PSDB; AB379587.

XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
PT

XX Disclosure; Page 460; 815pp; English.

XX The present invention relates to proteins from *Neisseria gonorrhoeae*.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid

CC molecules of the invention.

XX Sequence 1015 AA;

XX Query Match 6.9%; Score 105; DB 24; Length 1015;

XX Best Local Similarity 21.5%; Pred. No. 8;

XX Matches 74; Conservative 55; Mismatches 113; Indels 102; Gaps 17;

QY 9 TVEVKSALTADGALDLYNKYLDVLPWKT-----FDEIKELSRFKQYEQEASVLVG 63

DB 104 TADPAKSVLDTAD-AMQSINSQVRQVTSSETEYLAVQQQLDTPANTRASLESTANLYVS 162

QY 64 DIKVL-----MDSQD--KYFEATQTVYEWCGV-----VTQLLSAYILLDFDEYNEK 107

DB 163 TSRAKLDYGYTQBEILKFTTEANAMTIGVGQAQQAALMQLSGVLQGDFF--K 220

QY 108 KASAQDKILIRIL-----DDGVKKL-NEAQ-----XSLTSSQSFNNASGKLLALD 152

DB 221 SISEAAPILLDTAAYNGKSRDEIKLSEGLTADVIFKAISGASEKFGQQAQKPVMTM 280

QY 153 SQLTNDFSEKSYFQSQVDRIKE-----AVAGAAA- 183

DB 281 GQALTVEFNN---WQSMVSKLLNDSCTMGIAAVIKLIADNLNLVPIVAGFAVAAAV 337

QY 184 -----GIVAPFGIIISYIAAGVIEGKLIPELNNRLKTVQNFSTLSATVKQANKD 235

DB 338 APTLALNALLANPFGII-----AVAIGTWG-LIAKFGDEIDVFGGWSNLS-----D 385

QY 236 IDAALKLATEIA--AIGEIKTETETTRFVVDYDMLSLKGA 277

DB 386 VIRAVWIIITVGEAVGTVKS-----FDGLTGRNLNEGA 420

RESULT 32

ABP30285

ID ABP30285 standard; Protein; 318 AA.

XX ABP30285;

XX DT 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 9746.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tettelin H;

XX WPI: 2002-352536/38.

DR N-PSDB; ABN70916.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

XX Claim 1; Page 4101; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC Streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

XX Sequence 318 AA;

XX Query Match 6.9%; Score 104.5; DB 23; Length 318;

XX Best Local Similarity 19.2%; Pred. No. 2;

XX Matches 55; Conservative 51; Mismatches 115; Indels 65; Gaps 11;

QY 27 YNKYLDQVLPKTFDEIKELSRFKQYEQEASVLVGDIKVLNDSQDKYFEATQTVYEW 86

DB 55 YHLYSLACPWASRVLMRKLN-----ESHISIVNPLML-----ENGWTFQCY 101

QY 87 CGVTVQLLSAYILLDFDEYNEKKAQKDILIRILDGVKKL-----NEAQSLTSSQSF 141

DB 102 KGVIPDMINOSQYLYQYQASQSDYTGRTVTVPLWD--KKFHTIVNSESSEIMELMTAF 159

QY 142 NNASGKLLALDSQTNDFSEKSYFQSQVDRIKEAYAGAAAGIVAGPFGIIISYSTAAG 201

DB 160 NHITGN-----TDDYTPDS--LQGDIMNNFIYP-----KINNG 192

QY 202 VIEGKLIPELNNRLKTVQNFSTLSATVKQANKDIDAAKLKLAIEAIGIKETETTR 261

DB 193 VKAGFATSQNVYQKEVETLFTAL-----DQLEKLSDNHVLVGEQFTEADIRL 241

QY 262 F--YVDYDMLSLKGAAGKQVINTCNEYQQR--HGKKTLPFVDPVA 304

DB 242 FTTLVRFD---TVVYGHFKCNKALHDYPLHMYTKRIYNLPGLIA 283

RESULT 33

ABP28740

ID ABP28740 standard; Protein; 339 AA.

XX ABP28740;

XX DT 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 6656.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus agalactiae.

XX WO200234771-A2.

QY 66 KVLNDSQ---DKYFEATQTVWCVVTVQLLSAYILLFDEY----- 104
 Db 252 VIKLNDKMPEDIDKFSVLTLQEKIPEIQNAGKQISMDDDFASVETMTQGIQEAQGL 311
 QY 105 -----NEKKASAKOILIRILDDGVKKLNEAQKSL-----LTSQSFFNNSG 146
 Db 312 QIINQVQKSMPIKGLGDADSLGTITLDAAKMQSALPSITNSVQITLQIQFSKNTS 371
 QY 147 KLAL-----DSQLTNPFSEK-SSYFOSQVDRIRKEAYAGAGIVAGPGLLIISVIA 199
 Db 372 SVISVIDQALADNQLTDDKEQINSLTNFTNNAKQREA----- 411
 QY 200 AGVIEGKLIPELNNRLKTVQ--NFFTSLSATVKQANKDIDAALKLATEIAAIGIKETET 257
 Db 412 -----IQNIVEYMKVQESNGNHDLSITIEQLT-NLDSLSLSLSTRMSHLNDLVQEG 462
 QY 258 ETT--RFVY-----DYDDL-----MLSLKGAAKWMINT-----CNEYQQ 290
 Db 463 DVTKIRAVLSQINDVYNNISDLINKIDVSEISSTINKALTTLTINTIDAKGLNQAQQ 520

 RESULT 37
 ABG31849
 ID ABG31849 standard; Protein; 2334 AA.
 AC ABG31849;
 DT 05-NOV-2002 (first entry)
 DE Human kinase, MEK1.
 XX
 KW Extracellular signal regulated kinase; hyperalgesia; surgery;
 KW opoid withdrawal; pain sensitisation; analgesic; chronic pain;
 KW ERK; MEK1; human; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN WO200258687-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 25-JAN-2002; 2002WO-US02128.
 XX
 PR 25-JAN-2001; 2001US-264336P.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Gutstein HB;
 XX
 DR WPI; 2002-608414/65.
 DR N-PSDB; ABX90804.
 XX
 PT Reducing or reversing tolerance, physical dependence, hyperalgesia,
 PT withdrawal symptoms, or pain sensitisation in patients on analgesics
 PT for chronic pain, comprises inhibition of the extracellular
 PT signal-regulated kinase (ERK) -
 XX
 PS Disclosure; Page 156-161; 163pp; English.
 XX
 CC The invention relates to a method of reducing or reversing tolerance,
 CC reducing the risk of physical dependence or hyperalgesia, reducing the
 CC symptoms of opoid withdrawal or inhibiting pain sensitisation in a
 CC patient taking analgesics. The method comprises administering an
 CC analgesic and an extracellular signal-regulated kinase (ERK) inhibitor
 CC comprised in a formulation to reduce or reverse tolerance, risk
 CC of physical dependence, hyperalgesia, symptoms of opoid withdrawal, or
 CC inhibiting pain sensitisation in patients taking analgesics for chronic
 CC pain or those undergoing surgery. The present sequence represents
 CC the amino acid sequence of human MEK1 (not defined).
 XX
 SQ Sequence 2334 AA;

Query Match 6.9%; Score 104; DB 23; Length 2334;
 Best Local Similarity 25.6%; Pred. No. 28;
 Matches 60; Conservative 39; Mismatches 85; Indels 50; Gaps 13;

 QY 78 EAFQTV-YWCVGVTVQLLSAYILLFDEYNEKK-ASAQKILIRILD---DGVKKLNEAQK 132
 Db 1203 KATESYQYKDGKGVNTSKDAYGTETVEYNNKNDVTWKMDTEGNTVDIAYDGLDAVSETDQ 1262
 QY 133 SLTSS-----QSNN--ASKLLALDSQLTNDPS---EKSSY-FOSQVDRIRKEAYAGAA 182
 Db 1263 SGKSSAAVYDKGNQIQSKKSLASINILKQSGFAKSGWNLTRASKDR-RKISVIADK 1321
 QY 183 AGIVAGPFGI-ITSYIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQANKDIDAALK 241
 Db 1322 SGVLSGSKALEVLSQTSAGTDHG-----YSSATQTV-----EL 1355
 QY 242 KLATEIAAIGIKETETETTFYVDYDDLMLSLKGAAKWMINTC-NEYQQRHGK 294
 Db 1356 EPNTTYSGLTKITDLAKRAYFNID-----LRDKQKRIOWIHNEYSALAGK 1403

 RESULT 38
 ABJ10604
 ID ABJ10604 standard; Protein; 1959 AA.
 AC ABJ10604;
 DT 28-NOV-2002 (first entry)
 DE Human novel protein NOV11 SEQ ID NO: 36.
 XX
 KW Human; NOVX; single nucleotide polymorphism; SNP; anti-HIV, cytostatic;
 KW antiarteriosclerotic; antidiabetic; antiaesthetic; antiinflammatory;
 KW antidiabetic; hypotensive; neuroprotective; anorectic; nootropic;
 KW antidepressant; immunosuppressive; antibacterial; antiparasitic;
 KW virucide; tranquilizer; anticonvulsant; osteopathic; analgesic;
 KW antiparkinsonian; dermatological; antiinfertility; cerebroprotective;
 KW antiaddictive.
 XX
 OS Homo sapiens.
 XX
 PN WO200259315-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 19-DEC-2001; 2001WO-US50076.
 XX
 PR 19-DEC-2000; 2000US-256619P.
 PR 19-JAN-2001; 2001US-262959P.
 PR 28-FEB-2001; 2001US-272408P.
 PR 20-APR-2001; 2001US-285189P.
 PR 26-JUL-2001; 2001US-308039P.
 PR 09-AUG-2001; 2001US-311266P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Patturajan M, Vernet CAM, Casman SJ, Malyankar U;
 PI Shenoy S, Spytek KA, Gangolli E, Miller C, Boldog F, Li L;
 PI Taupier RJ, Kekuda R, Smithson G, Zerhusen BD, Liu X, Colman SD;
 PI Tchernev V, Si J, Edinger S, Stone D, Sciore P, Millet I;
 PI Rothenberg M;
 XX
 DR WPI; 2002-666903/71.
 DR N-PSDB; ABT08505.
 XX
 PT New isolated NOVX polypeptides and polynucleotides, useful for
 PT preventing, diagnosing or treating NOVX-associated disorders e.g.
 PT diabetes, Crohn's disease, atherosclerosis, cancer, Huntington's
 PT disease or Alzheimer's disease -
 XX
 PS Claim 1; Page 138; 363pp; English.
 XX
 CC The present invention provides the protein and coding sequences of

CC several novel human proteins, designated NOVX. These can be used in the
CC treatment of diseases such as cancers, Hodgkin's disease, Von
CC Hippel-Lindau syndrome, Alzheimer's disease, stroke, tuberculous sclerosis,
CC hypercalcaemia, Parkinson's disease, Huntington's disease, cerebral
CC palsy, epilepsy, leuko-lymph syndrome, multiple sclerosis, ataxia
CC telangiectasia, leukodystrophies, addiction, anxiety, depression, pain,
CC obesity, Crohn's disease, osteoporosis, inflammatory bowel disease,
CC infertility, atherosclerosis, hypertension, scleroderma, haemophilia,
CC diabetes, pancreatitis, autoimmune disease, asthma, arthritis,
CC immunodeficiencies, HIV, viral, bacterial or parasitic infections, or
CC graft-versus-host disease. The present sequence is a protein of the
CC invention.

XX
SQ Sequence 1959 AA;
Query Match 6.8%; Score 103.5; DB 23; Length 1959;
Best Local Similarity 19.2%; Pred. No. 25;
Matches 68; Conservative 49; Mismatches 119; Indels 119; Gaps 13;

QY 7 EQTVVVKSAIE---TAGALDLYNKYLQVLPKTFDETIKE-----LSRPKQYS 55
Db 1166 EQEVNLIKKTLEBEAKTHEAQIQEMRQKHSQAV--BELAEQLEQTKRVKANLEKAKQTLE 1223
QY 56 QEASVLVGDIKVLLMDSQKYPATQTVYEWCGVTTQLLSAVILLFDEYNEKKAQAQDI 115
Db 1224 NERGELANEVKULLQGRD-----SEKRRKKVZAQLE 1256
QY 116 LIRILDGVKUNEAQKSLTSSQPNNASGKLLALD---SOLTNDPSEKSYFOSQVDR 172
Db 1257 LQKFNVEGRVRLTADKVTQLQVELDVTGLSQSDSKSLTKDFSALESQLODTQEL 1316
QY 173 IRKEAVAGAAAGVAGPFGLLISYSIAAGVIEGKLPELNNRLKTVO---NPF----- 222
Db 1317 LQEN-----RQKL--SLSTKLQVEDEKNSFRQLSEE 1348
QY 223 -----TSLSATV-----KOANKDIDAACKLATAIAIG 251
Db 1349 EAKHNLKQIATLHAQVADMKKQWDSVGCLETAEBVRKLOKLEGLSQRHEEKVAAVD 1408
QY 252 EIKETETTFYDYDDMLSLKGAQKXWNTCN-EYQORHGKTKLEVPDVAS 305
Db 1409 --KLETKTKLQELDLDLVL-----DHQSQACNLEKQKQKPDQLLAEKTIISA 1457

RESULT 39
ABBS9344
ID ABB59344 standard; protein; 2056 AA.
XX
AC ABB59344;
XX
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 4824.
XX
XX Drosophila melanogaster polypeptide; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.

DR N-PSDB; ABL03447.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 4824; 21bp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU1840-ABU16175) and the encoded proteins
CC (ABBS7737-ABBS7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 2056 AA;
Query Match 6.8%; Score 103.5; DB 22; Length 2056;
Best Local Similarity 19.9%; Pred. No. 27;
Matches 74; Conservative 53; Mismatches 112; Indels 133; Gaps 15;
QY 8 QTVVVKSAIETADGALDLYNKYL-----DQVLPKTFDETIK 45
Db 1296 ENLRKAKTVLEKAGTLEAENADLATELSVNSROENRRRKQAESQIAELQV---KUA 1352
QY 46 ELGRFKQYEQEASVLVDIKVLLMDSQKYPATQTVYEWCGVTTQLLSAYILLFDEYN 105
Db 1353 EIERARSELQKCTKQEAENITNQLAEALKASAAVKSASNNESQLEAQQLLEETR 1412
QY 106 EK-----KASAKDI-LIRI 119
Db 1413 QKGLSKLRQIESEKALQEQLEEDDEAKRNTVERKLAETVTTOMQIKKAEEDADLAE 1472
QY 120 LDDGVKLLN-----EAQ-KSLTSSQSFNNASGKL-----LALDSQLTNDFS-EKS 163
Db 1473 LSECKKLNKDIEALEEQVKELIAQNDRLDKSKKIQSELEDATEIEAQRTKVLEKK 1532
QY 164 SYFOSQVDRTRKEAYAGAAAGVAGPFGLLISVIA-----AGVIEGKL----- 207
Db 1533 ---QKFNFKILAEKA-----ISEQIAQERDTAERAREKETRVLSVSREL 1575
QY 208 -----IPELNNRLKTVQNFFTSLSATVVKQANKDI-DAACKLATAIAAIGKIKTETTT 260
Db 1576 DEAFDKIEDLENKKTQLQNELDDLANTQGTADKNVHELEKAKALE-SQLAELKAQNEEL 1634
QY 261 RFYVDYDMLMS 272
Db 1635 E-----BDLQUT 1641

RESULT 40
AAAY9936
ID AAY49936 standard; protein; 2101 AA.
XX
XX AAY49936;
XX
XX 01-FEB-2000 (first entry)
XX
XX Human Numa protein #1.
XX
XX Human; Numa; intranuclear protein; mitosis; segregation; DNA-PK-cs;
KW DNA-activated protein kinase; catalytic subunit; PARF; detection;
KW nuclear NAD+ ADP-ribosyltransferase; autoantigen; immune response;
KW autoimmune disease; cancer; type I diabetes mellitus; thyroiditis;
KW myasthenia gravis; primary biliary cirrhosis; rheumatoid arthritis;
KW systemic lupus erythematosus; polymyositis; dermatomyositis;
KW Sjogren's syndrome; scleroderma; Graft-vs-host disease.
XX

XX SQ Sequence 606 AA;
 Query Match 6.8%; Score 103; DB 17; Length 606;
 Best Local Similarity 19.8%; Pred. No. 6;
 Matches 66; Conservative 69; Mismatches 131; Indels 68; Gaps 17;
 QY 1 MTSFAEQTVVVKSAIETADGALDLYNKYLQVTPKTFDETIKELSRFKQYEQEASV 60
 DB 228 LQNLLEKEVELEKHAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 284
 QY 61 LVGDIKVLNDSQDKYFEATQTVYEWCVVLTLLSAYILLDFEYNEKASAKQDILIRIL 120
 DB 285 L-RDVTQAQLESEQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSRDVTQAQ-L 334
 QY 121 DDGVKLNKAQSL---LTSQSFNNAAGKLLALDSQLTN-DFSEKSSVFQSOVDRIKE 176
 DB 335 ESQVEKYNDTAQSLRDVTAQLESYKSTLKEIE-DLKENLTQEKVMAEKSVEDVQOQ 393
 QY 177 AYAGAAAGIVAGPFGHIIISIAAGVIEGKLIPELNNR-----LKTQVQNF---TSL 225
 DB 394 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETTSFLEKITDL 436
 QY 226 SATVQKANKDIDAAKLATEIAAIGIKTE--TETRFYVDYDDL----- 269
 DB 437 KNQLRQODEDFRKQLEKKGKRTAEKENVMTLMEINKWLLYDELYEKTPTFQOQLDAF 496
 QY 270 ---MLSLK--GAAKWMINTCNE-YQORHGKKTLL 297
 DB 497 EAEKQALLNEHGATQEQOLNKIRDSYAQLLGHQNL 530
 RESULT 42
 AAR99675
 ID AAR99675 standard; Protein; 631 AA.
 AC AAR99675;
 XX AAR99675;
 DT 10-OCT-1996 (first entry)
 DE RHAMM 1-2a isoform.
 XX RHAMM 1-2a; receptor for hyaluronic acid mediated motility;
 KW hyaluronan receptor; cell locomotion; cell proliferation;
 KM breast cancer; therapy.
 XX Mus sp.
 XX Key Location/Qualifiers
 FT Region 55..79
 FT /note= "exon 2A-encoded region"
 XX EP721012-A2.
 XX 10-JUL-1996.
 XX 16-OCT-1995; 9SEP-0307310.
 XX 14-OCT-1994; 94GB-0020740.
 XX (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.
 XX (UTMA-) UNIV MANITOBA.
 XX Entwistle J, Turley EA, Zhang S;
 XX WPI; 1996-310997/32.
 XX N-PSDB; AAT34525.
 XX Receptor for hyaluronic acid-mediated motility protein, and DNA
 PT encoding it - useful to treat or prevent diseases associated with
 FT the receptor, e.g. breast cancer
 XX Claim 8; Page 50-52; 117pp; English.

XX RHAMM 1-2a (AAR99675) is an alternatively spliced variant of
 CC RHAMM 1 (AAR99673) (receptor for hyaluronic acid mediated
 CC motility), a protein involved in cell locomotion or motility and
 CC cell proliferation and transformation. It differs from RHAMM 1
 CC by an insertion of 25 amino acids (see also AAR99674) between
 CC amino acids 54 and 55 of RHAMM 1, resulting from an alternatively
 CC spliced exon 2A (AAT34502). RHAMM 1-2a is the isoform that is
 CC overexpressed in tumors. Determination of the level of RHAMM
 CC 1-2a in a sample can be used to assess the prognosis of a tumour
 CC (esp. breast cancer) patient. The RHAMM 1-2a protein can also
 CC be used to suppress or control a tumour by modulating the
 CC interaction of cell-associated RHAMM with its ligand.
 XX Sequence 631 AA;
 Query Match 6.8%; Score 103; DB 17; Length 631;
 Best Local Similarity 19.8%; Pred. No. 6.3; Indels 68; Gaps 17;
 Matches 66; Conservative 69; Mismatches 131;
 QY 1 MTSFAEQTVVVKSAIETADGALDLYNKYLQVTPKTFDETIKELSRFKQYEQEASV 60
 DB 253 LQNLLEKEVELEKHAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 309
 QY 61 LVGDIKVLNDSQDKYFEATQTVYEWCVVLTLLSAYILLDFEYNEKASAKQDILIRIL 120
 DB 310 L-RDVTQAQLESEQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSRDVTQAQ-L 359
 QY 121 DDGVKLNKAQSL---LTSQSFNNAAGKLLALDSQLTN-DFSEKSSVFQSOVDRIKE 176
 DB 360 ESQVEKYNDTAQSLRDVTAQLESYKSTLKEIE-DLKENLTQEKVMAEKSVEDVQOQ 418
 QY 177 AYAGAAAGIVAGPFGHIIISIAAGVIEGKLIPELNNR-----LKTQVQNF---TSL 225
 DB 419 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETTSFLEKITDL 461
 QY 226 SATVQKANKDIDAAKLATEIAAIGIKTE--TETRFYVDYDDL----- 269
 DB 462 KNQLRQODEDFRKQLEKKGKRTAEKENVMTLMEINKWLLYDELYEKTPTFQOQLDAF 521
 QY 270 ---MLSLK--GAAKWMINTCNE-YQORHGKKTLL 297
 DB 522 EAEKQALLNEHGATQEQOLNKIRDSYAQLLGHQNL 555
 RESULT 43
 ABB92030
 ID ABB92030 standard; Protein; 796 AA.
 AC ABB92030;
 XX ABB92030;
 DT 31-MAY-2002 (first entry)
 DE Herbicidally active polypeptide SEQ ID NO 1241.
 XX Herbicidal; plant; agriculture; herbicide.
 KW Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 PN WO200210210-A2.
 PD 07-FEB-2002.
 XX 28-AUG-2001; 2001WO-EP09892.
 XX 28-AUG-2001; 2001WO-EP09892.
 XX (FARB) BAYER AG.
 XX Tietjen K, Weidner M;
 XX WPI; 2002-269010/31.
 XX

PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT of plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -

XX Claim 5; SEQ ID NO 1241; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
 CC (AB90790-AB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 796 AA;

Query Match 6.8%; Score 103; DB 23; Length 796;
 Best Local Similarity 22.3%; Pred. No. 8.5;
 Matches 61; Conservative 50; Mismatches 124; Indels 38; Gaps 8;

QY 42 ETIKELSPKQYSEFA-----SVLVDIKVILMSQDKYFATQTVVWCGVTTLLSAY 97

DB 467 DKLKEQILKAKETSTEAEPSEVLNEMIEKLSKSEIDDEYTEAAIAV----GLEERLTA-- 520

QY 98 ILLFDEYNKKASAKQDILIRLDGKVKLNEAQKSLITSSQSFNNASGKLAL----- 151

DB 521 --MREEPS--KASSEHLMPVLIEKIEKLEENFRLTDAPNYESLKSLLNMLRDSRA 576

QY 152 --DSQLTNDFSEKSSYFOSQVDR-----IRKEAVAGAAAGIVAGPFLGIISYIAA 200

DB 577 KAASEATSLKKEINKRFQEAVDPRPEIREKVEAIKAEVASSGASGFFDELDPALKEKVLKTK 636

QY 201 GVIEGKL--IPELNNRLKTVQNPFTSUSATVKQAKDIDAALKLATEIAAIGEIKTET 257

DB 637 GEVEAEVAGLVKSGLELDVAKQKQKOTAEQIYAENLQEKLEKLNQETSKIEEVRT 696

QY 258 ETTTFYVDYDMLSLKGAAKKMTNCTNEYQQ 290

DB 697 PEIKSMVE---LLKVETAKASKTPGVTEAYQK 725

RESULT 44

AA93246

ID AAY93246 standard; Protein; 974 AA.

XX AC AAY93246;

XX DT 04-SEP-2000 (first entry)

XX DE An Escherichia coli virulence protein.

XX KW Virulence protein; tata; tatB; tatC; tatE; mdoG; creC; recG; yggN;

XX KW eck1; iroB; iroC; iroE; mtd2; msi; vaccine; infection;

XX KW Gram negative bacterium.

XX OS Escherichia coli.

XX PW WO200028038-A2.

XX PD 18-MAY-2000.

XX PF 09-NOV-1999; 99WO-GB03721.

XX PR 09-NOV-1998; 98GB-0024569.

XX PR 09-NOV-1998; 98GB-0024570.

XX PR 17-DEC-1998; 98GB-0027814.

XX PR 17-DEC-1998; 98GB-0027815.

XX PR 17-DEC-1998; 98GB-0027816.

XX PR 17-DEC-1998; 98GB-0027818.

XX PR 13-JAN-1999; 99GB-0000708.

PR 13-JAN-1999; 99GB-0000710.

PR 13-JAN-1999; 99GB-0000711.

PR 28-JAN-1999; 99GB-0001915.

XX (MICR-) MICROSCIENCE LTD.

XX Crooke HR, Clarke EE, Everest PH, Dougan G, Holden DW, Shea JE;

PI Feldman RG;

XX WPI; 2000-376550/32.

DR N-PSDB; AAA15186.

XX Peptide encoded by an operon including genes from Escherichia coli for

PT screening potential drugs, detecting virulence and treating conditions

PT associated with infection by a Gram negative bacterium -

XX Claim 2; Page 108-112; 122pp; English.

XX The present sequence represents an Escherichia coli virulence protein.

CC The specification describes virulence proteins which are encoded

CC by an operon including tata, tatB, tatC, tatE, mdoG, creC, recG, yggN,

CC eck1, iroD, iroC, iroE, mtd2 or msi-16 genes obtained from Escherichia

CC coli K1. The virulence proteins and polynucleotides, and their vaccines

CC are useful for screening potential drugs, for the detection of virulence,

CC and for treating or preventing conditions associated with infection by

CC a Gram negative bacterium particularly Escherichia coli.

XX Sequence 974 AA;

QY Query Match 6.8%; Score 102.5; DB 21; Length 974;

Best Local Similarity 19.5%; Pred. No. 12;

Matches 69; Conservative 53; Mismatches 95; Indels 137; Gaps 14;

QY 7 EOTVEVKAIAITADGALDLYNKYLDQVLPWKTFDETIK----ELSR--FKOEYSQEASV 60

DB 417 ELTKENNASLVEKAGREGFIENK-----PYQPKEMLENFIEIARFFKDD----- 463

QY 61 LVGDIKVLMSQDK-----YFEATQTVVWCGVTTQ 92

DB 464 --GDMSELFVETKQRRNEEDLLSKRSKQTKAKDKLKDLYDFDKLDNDY--WNIEINK 520

QY 93 LLSAVILFPD-----EYNEKASAKQDILIRLDGV-----KLN----- 128

DB 521 LINKNEEYFSSTEITDNIIDYVYVYKIKENDAIKLNLSVDIKKPSGVGLTKELSNLWD 580

QY 129 ----EAQKSLTSSQSFNNASGKLALDSQLTNDP-----SKSSYFQSQVDR 173

DB 581 RYOIERQKILLSNELKDNVDRKLELDNK--NDFNLKRLKLEDSLNLQSQSYKEKELTKL 639

QY 174 RKEAYAGAAAGIVAGPFLGIISYSIAAGVIEGKLPELNNRLKTVQNPFTSUSATVKQAN 233

DB 640 YNDA-----KNALKDVQS-----KAN 655

QY 234 KQIDAARKLATEIAAIGSIKTTETTFYVDYDMLSLKGAAKKMTNCTNE 287

DB 656 RLISDNKKHKSELK--NISVFQSTNLNGKDTAVILDVKNLESKIENTSNE 706

RESULT 45

AA49937

ID AAY49937 standard; protein; 2115 AA.

XX AC AAY49937;

XX DT 01-FEB-2000 (first entry)

XX DE Human NUMA protein #2.

XX KW Human; NUMA; intranuclear protein; mitosis; segregation; DNA-PK-cs;

XX KW DNA-activated protein kinase; catalytic subunit; PARP; detection;

XX KW nuclear NAD+ ADP-ribosyltransferase; autoantigen; immune response;

XX KW autoimmune disease; cancer; type I diabetes mellitus; thyroiditis;

XX KW myasthenia gravis; primary biliary cirrhosis; rheumatoid arthritis;

XX systemic lupus erythematosus; polymyositis; dermatomyositis;
KW Sjogren's syndrome; scleroderma; graft-vs-host disease.
XX Homo sapiens.
XX WO9953757-A1.
XX 28-OCT-1999.
XX 22-APR-1999; 99WO-US08774.
XX 22-APR-1998; 98US-0082643.
XX (MERI) MERCK & CO INC.
PA (UJO) UNIV JOHNS HOPKINS.
PA (MERI) MERCK FROST CANADA INC.
XX Thornberry N, Rosen A, Casciola-Rosen L, Andrade FA, Nicholson D;
PI Roy S;
PI WPI; 2000-013162/01.
XX Autoantigenic fragments useful for diagnosis, treatment of autoimmune
PT diseases and cancer -
XX Claim 4; Fig 10; 88pp; English.
XX The present invention describes autoantigenic fragments (I) produced by
CC the action of a lymphocyte granule enzyme (II) on isolated cells
CC containing an autoantigen. (I) is useful for prophylactic and therapeutic
CC treatment of an autoimmune disease (AI) such as type I diabetes mellitus,
CC thyroiditis, myasthenia gravis, primary biliary cirrhosis, systemic lupus
CC erythematosus, rheumatoid arthritis, polymyositis, dermatomyositis,
CC Sjogren's syndrome, scleroderma and graft-vs-host disease. (I) produced
CC by the action of (II) on cells isolated from target tissue is
CC administered to a patient for tolerising to the presence of (I). As a
CC therapeutic treatment, isolated (I) associated with AI condition is
CC contacted with serum of the patient containing autoantibodies (AAB)
CC against (I), under condition to allow the in vivo binding of AAB to (I)
CC and a portion of the AAB is removed from the serum of the patient. (I)
CC derived from malignant cell is administered to a patient for stimulating
CC an immune response against malignant cells to treat cancer. (I) is also
CC useful in assays for assessing the presence or absence of an AI condition
CC in a patient. Such assays involve detecting AAB in a sample by using (I)
CC or detecting (I) in a sample by using antibodies that specifically binds
CC to a cryptic epitope of (I). The present sequence represents the human
CC NUMA protein, which is a specifically claimed autoantigen from the
CC present invention.
XX Sequence 2115 AA;
SQ
Query Match 6.8%; Score 102.5; DB 21; Length 2115;
Best Local Similarity 22.3%; Pred. No. 33;
Matches 65; Conservative 44; Mismatches 109; Indels 73; Gaps 13;
QY 42 ETIKESLRKQESYQ-----EASVLVGDIKVLMD-----SOPKEATQTVYEWCG 88
DB 294 ETLKQCDLTKERSQMDRKINQSENGDLSPKREFASHLQLODALNELTEHSKAC- 352
QY 89 VVQLLSAYLLDFEYNEKASQKQDILRLDLDGVKLNKAEKXSLTSSQFNWASGKL 148
DB 353 -----QEWLEKQAEKELSALQD---KKCEKNEI-----QGLK 387
QY 149 LALDSQIT-----NDFSEKSSYFQSDVRIRKAYAGAAAGIVAGPGL-----IISYSIAAG 201
DB 388 SQLEEHLSQLODNPQKGBVLG---DVLQLEFLKGEATLAANNQLQARVEMLETERG 444
QY 202 VIEGKLIPE-----LNNRLKTVQNFSTLSATVQVQAKNDIDAKLKLATIAA-I 250
DB 445 QQEAKLAERHFEFEKQKLSLITDQSSISNLS-----QAKELEQASQAHGALTAQV 500
QY 251 GEIKTETETTRFVVDYDMLMLLKGAAK-----KMINTCNEYQQ-----RH 292

Db 501 ASLTSELTTLNATIQDQDELAKGQAKKEKQQLAQTQQQEQASQGLRH 551
RESULT 46
ABP73939
XX ABP73939 standard; Protein; 1038 AA.
AC ABP73939;
XX 30-JAN-2003 (first entry)
XX Candida albicans essential protein SEQ ID NO 7776.
XX Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX Candida albicans.
XX OS
XX WO200253728-A2.
XX 11-JUL-2002.
XX 26-DEC-2001; 2001WO-US49486.
XX 29-DEC-2000; 2000US-259128P.
PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX (ELIT-) ELITRA PHARM INC.
PA Rosner T, Jiang B, Boone C, Bussey H, Ohlsen KL;
PI WPI; 2002-566694/60.
XX N-PSDB; ABZ32489.
XX Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
PT expression -
XX Claim 44; SEQ ID NO 7776; 167pp + Sequence Listing; English.
XX The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthesis, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX Sequence 1038 AA;
Query Match 6.7%; Score 102; DB 23; Length 1038;
Best Local Similarity 20.8%; Pred. No. 15;

Matches 64; Conservative 59; Mismatches 118; Indels 66; Gaps 13;

QY 7 EQTEVVKVKS---ALETADGALDLYNKYLDQVTPKKTDETIKELSRKQB-YQSBASVLV 62
 Db 237 EETQALKSENEITAKVSELEDYKHS-----VEFDVVMKQNDQFQERHELEAAIDT 291
 QY 63 GDIVLLMDSQDKYFEATQTVYEWGVVTVQLLSAYILLFDEYNEKKAQKDLIRILD- 121
 Db 292 LHQTEATIQQSQRENTE-----LQQLKSTELDKQCEMNRLLASKNENLEMDLSEK 344
 QY 122 -DGVKLNEAKQSILTSQSFNNASGKLLALDSQLTN--DPSEK-SYFQSGVDRIK-KE 176
 Db 345 TDNLKELN---NKVLSQAQENLLETLDLTNSQFENNTDGNKLMKQNLSESLQNKVQTQE 401
 QY 177 AYAGAAAGIVAGPFGELIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQAKDI 236
 Db 402 AF-----IDELHHEQKIDNEY-----KAKIKDL 425
 QY 237 DAALKLATEATAAIGEIKTETETTRFYVDYDML--SLKGAAGKGMINTCNEYQQR--- 291
 Db 426 EYENALQSEETISRIRAKNSQYDDBAQAHOYEDQLKQENALQNDVVKYLNFKELDKKEVE 485
 QY 292 HGKKTLP 298
 Db 486 HAHQIAP 492

RESULT 47
 AAM39213
 ID AAM39213 standard; Protein; 1453 AA.
 AC AAM39213;
 XX
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 2358.
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 OS Homo sapiens.
 XX
 FN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0582317.
 PR 09-JUL-2000; 2000US-0596042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI58369.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX

PS Example 4; SEQ ID NO 2358; 10078bp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX SQ Sequence 1453 AA;

Query Match 6.7%; Score 102; DB 22; Length 1453;
 Best Local Similarity 21.1%; Pred. No. 22;
 Matches 56; Conservative 53; Mismatches 88; Indels 68; Gaps 10;

QY 52 QEYSQESASVLGDIKVLMDSQDKYFEATQTVYEWGVVTVQLLSAYILLFDEYNEKKASA 111
 Db 972 EKVTAAEKIKKMEIEILLLEDQNSRFIEKKLME-----DRAECSSQLAESEKAKNLA 1026
 QY 112 ---OKDILIRILDGKVLKNEAKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYFQ 167
 Db 1027 KIRNQEWISDLBSRLKKEKTRQLEKAKRK-----LDGE-TTDLQDQIAELQ 1075
 QY 168 SQVDRI-----RKEAYAGAAAGIVAGPFGELIISYSIAAGVIEG-----KLPELNNR 214
 Db 1076 AQIDELKLQAKKEEELQA-----LARGDDETLHKNNALKVVRLEQAQ 1119
 QY 215 LKTVQNFFTSLSATVKQAKDIDAALKLATEIAAIGEIKTETETTRFYVDYDMLSL 274
 Db 1120 IAEQLDEFESEKASRNKAEKQ-----KRLDSELEA---LKTELEDT-----LD 1160
 QY 275 KGAAKGMINTCNEYQQRHGKKTLP 299
 Db 1161 TTAAQQLRTKREQEVAEKKALEE 1185

RESULT 48
 AAM39214
 ID AAM39214 standard; Protein; 1469 AA.
 XX
 AC AAM39214;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2359.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 OS Homo sapiens.
 XX
 FN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR

PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR N-PSDB; AA158370.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2359; 10078bp; English.
PS
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA138642-AA142213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 1469 AA;
SQ
Query Match 6.7%; Score 102; DB 22; Length 1469;
Best Local Similarity 21.1%; Pred. No. 23;
Matches 56; Conservative 53; Mismatches 88; Indels 68; Gaps 10;
Qy 52 QEYSQASVLVDGDKVLLMDSQDKYFEATQTVVEMCGVVTQLLSAVILLDFEYNEKKASA 111
Db 988 EKVTAEAKIKWEEIILLDDQNSKFKTEKKLME-----DRIACSSQLAEKEKAKNLA 1042
Qy 112 ----QKDLIRILDGKVLKNEAKSLTSSQSFNNASGKLLALDSQLTNDPSEKSSYQ 167
Db 1043 KIRNKQEVMSDLEERLKEKTEKQLEKAKRK-----LDGE-FTDLDQDIAELQ 1091
Qy 168 SQVDRI-----RKEAVAGAAAGIVAGPFGILISYSIAAGVIEG-----KLIPELNLR 214
Db 1092 AQIDELKQLAKKEELQGR-----LARGDDETHKNNAKVVRELQ 1135
Qy 215 LKTVQNFTSLSATVQKANDIDAKUKLATEIAAIGETETETTFYVDYDMLSL 274
Db 1136 IAEQLQDFEFSKARNAEKQ-----KRDLSLEEAL-----LXTELEDT-----LD 1176
Qy 275 KGAAKMINTCNEQVRHGKKTLEF 299
Db 1177 TTRAQQLRTKKEQEVAEKKALEE 1201
RESULT 49
ABU07447
ID ABU07447 standard; Protein; 1976 AA.
XX
XX AC ABU07447;
XX
XX 28-JAN-2003 (first entry)
XX

DE Protein differentially regulated in prostate cancer #50.
XX
XX Prostate cancer; gene expression; differential regulation;
KW molecular marker; drug target; cancer detection; cancer diagnosis;
KW cancer staging; cancer grading; cancer assessing; cancer monitoring.
XX
XX Homo sapiens.
OS
XX WO200281638-A2.
PN
XX 17-OCT-2002.
PD
XX
XX 08-APR-2002; 2002WO-US10924.
PF
XX 06-APR-2001; 2001US-281731P.
PR
XX 06-APR-2001; 2001US-281732P.
PR
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
PA
XX
XX Sun Z, Jay G;
PI
XX WPI; 2003-058520/05.
DR N-PSDB; ABX10349.
XX
XX Novel genes which are differentially regulated in prostate cancer.
PT useful for diagnosing prostate cancer in prostate tissue sample and
PT assessing therapeutic or preventive intervention in prostate cancer
PT patients -
XX
XX Claim 1; Page 309-315; 416pp; English.
PS
XX The invention describes genes (I) which are differentially regulated in
CC prostate cancer. (I) is useful for diagnosing a prostate cancer in a
CC sample comprising prostate tissue, which involves determining the number
CC of target genes which are differentially-regulated in the sample, where
CC the number is indicative of the probability that the sample comprises
CC prostate cancer. (I) is useful for assessing a therapeutic or preventive
CC intervention in a subject having a prostate cancer, which involves
CC determining the expression levels in a sample comprising prostate tissue
CC of target genes which are differentially-regulated in prostate cancer.
CC Preferably, the expression levels of at least 10 genes are determined.
CC (I) is also useful for identifying agents that modulate a biological
CC activity of a polypeptide differentially-regulated in prostate cancer
CC cells, which involves contacting a polypeptide differentially-regulated
CC in prostate cancer cells with a test agent under conditions effective for
CC the test agent to modulate a biological activity of the polypeptide, and
CC determining whether the test agent modulates the biological activity.
CC (I) is useful as molecular markers, as drug targets, and for detecting,
CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC preventing or treating, determining predisposition to diseases and
CC conditions especially relating to prostate cancer. (I) and its expression
CC products are used in the diagnostic test to assay for presence of cancer
CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC blood etc. (I) is useful for assessing cancer e.g., to determine the type
CC of cancer, its stage of development, the nature of genetic defect, etc.
CC The polypeptide encoded by (I) can be used as target for therapy or drug
CC discovery. (I) can also be used for expressing the polypeptide and thus
CC for searching specific binding partners of the polypeptide. (I) is
CC useful in therapeutic applications to treat prostate cancer. The
CC identification of specific genes, and groups of genes, expressed in
CC pathways physiologically relevant to prostate cancer permits the
CC definition of functional and disease pathways and the delineation of
CC targets in these pathways which are useful in diagnostic, therapeutic,
CC and clinical applications. This is the amino acid sequence of a protein
CC differentially regulated in prostate cancer.
XX
XX Sequence 1976 AA;
SQ
Query Match 6.7%; Score 102; DB 24; Length 1976;
Best Local Similarity 21.1%; Pred. No. 33;
Matches 56; Conservative 53; Mismatches 88; Indels 68; Gaps 10;
Qy 52 QEYSQASVLVDGDKVLLMDSQDKYFEATQTVVEMCGVVTQLLSAVILLDFEYNEKKASA 111

Db 972 EKVTAIAKIKWEEBELLLEDDQNSKFIKEKLM-----DRIACSSQLAESEKAKNLA 1026
QY 112 ----OKDILIRLDGVKVLNEAOKSLTSSQSPNNSGKLLALDSQTNDFSEKSSYFQ 167
Db 1027 KIRNKQEVMSIDLERIKKESKTRQELEKAKRK-----LDGE-TTDLQDQIAELQ 1075
QY 168 SQVDRI-----RKEAYAGAAAGIVAGPFGLIISYSIAAGVIEG-----KLIPELNNR 214
Db 1076 AQIDELKQLAKKEEELQGA-----LARGDDETLHKNNALKVVELOAQ 1119
QY 215 LKTQVNFSTLSATVKQANKIDIDAKKLATEIAAIGIKETETTRFYVDYDMLSL 274
Db 1120 IAELOEDFESEKASRNKAEKQ-----KRDLSLEEA---LKTELEDT-----LD 1160
QY 275 KGAAKKMINTCNEYQQRHGKKTLE 299
Db 1161 TTAQQLRTKREQVAVELKALBE 1185

RESULT 50

AAM40999

ID AAM40999 standard; Protein; 1988 AA.

XX AAM40999;

XX AC AAM40999;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 5930.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX KW peripheral nervous system; neuropathy; central nervous system; CNS;

XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX KW leukaemia.

XX OS Homo sapiens.

XX FN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0498725.

XX PR 25-APR-2000; 2000US-052317.

XX PR 09-JUL-2000; 2000US-0596042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Dmanac RT;

XX DR WPI: 2001-442253/47.

XX DR N-PSDB; AAI60155.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Example 2; SEQ ID NO 5930; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,

XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful

XX CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activity, such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX SQ Sequence 1988 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 6.7%; Score 102; DB 22; Length 1988;

Best Local Similarity 21.1%; Pred. No. 34;

Matches 56; Conservative 53; Mismatches 88; Indels 68; Gaps 10;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Search completed: November 28, 2003, 13:51:14

Job time: 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2003, 13:48:16 ; Search time 20 Seconds

(without alignments)

1466.572 Million cell updates/sec

Title: US-09-993-292A-2

Perfect score: 1515

Sequence: 1 MTSIFAEQTVVEVKSALFET.....NEYQQRHGKTLFEVDPVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR 76.*

1: Piri.*

2: Piri2.*

3: Piri3.*

4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1503	99.2	305	2 AE0673	hemolysin HlyE [i
2	1402	92.5	305	2 C64864	hemolysin E - Esch
3	1387	91.6	305	2 E90838	hemolysin E [impor
4	1387	91.6	305	2 E85896	probable pore form
5	118	7.8	1496	2 T05634	hypothetical prote
6	116	7.7	1023	1 LEECA	hemolysin A - Esch
7	115.5	7.6	622	2 T22716	hypothetical prote
8	115	7.6	587	2 A90394	bP52 protein homol
9	114.5	7.6	478	2 T12818	hypothetical prote
10	114.5	7.6	652	2 B59102	hypothetical prote
11	113.5	7.5	127	2 AC1814	hypothetical prote
12	113.5	7.5	1127	2 T28317	ORF MSV156 hypotne
13	113.5	7.5	2823	2 T23064	hypothetical prote
14	113.5	7.5	2823	2 F87908	protein T2A3.8 [i
15	113.5	7.5	3102	2 T43291	laminin alpha chai
16	113	7.5	956	2 S30834	hypothetical prote
17	113	7.5	1005	2 A64855	hypothetical prote
18	112.5	7.4	1024	2 S10556	hemolysin A - Esch
19	112	7.4	1039	2 S18199	myosin heavy chain
20	111.5	7.4	1999	1 S21801	myosin heavy chain
21	111.5	7.4	2819	2 A90351	conserved hypotet
22	111	7.3	1295	2 T24587	hypothetical prote
23	110.5	7.3	584	2 S75986	hypothetical prote
24	110.5	7.3	927	2 AG1739	transmembrane prot
25	110	7.3	821	2 S67087	hypothetical prote
26	110	7.3	1938	1 NWKWL	myosin heavy chain
27	109.5	7.2	1492	2 T14652	protein J - Versin
28	109.5	7.2	1545	2 T14966	phage lambda-relat
29	109	7.2	595	2 F75008	hypothetical prote

30	109	7.2	739	2 H75001	methyl-accepting c
31	109	7.2	955	1 A35254	leukotoxin A - Fas
32	109	7.2	1098	2 B70232	hypothetical prote
33	108.5	7.2	1066	1 A48669	kinesin-related pr
34	108	7.1	726	2 T44825	hypothetical prote
35	108	7.1	998	2 T00237	hemolysin A toxin
36	108	7.1	2139	2 T18296	myosin heavy chain
37	107.5	7.1	1292	2 D84727	probable RAP50 DNA
38	107.5	7.1	1875	2 S38173	myosin-like protei
39	106.5	7.0	539	2 F72288	methyl-accepting c
40	106.5	7.0	927	2 AH1369	transmembrane prot
41	106.5	7.0	1272	2 C90593	hypothetical prote
42	106.5	7.0	1473	2 A35186	salivary agglutini
43	106.5	7.0	2346	2 T13829	Trp homolog - frui
44	106	7.0	998	2 I41078	hemolysin - Escher
45	105.5	7.0	540	2 T44967	gas-vesicle protei
46	105	6.9	1601	2 AB1730	hypothetical prote
47	104.5	6.9	566	2 S40591	hypothetical prote
48	104.5	6.9	1229	2 T48959	kinesin-like prote
49	104	6.9	329	1 VNU721	VSG expression sit
50	104	6.9	633	2 T41332	casp homolog - fis
51	104	6.9	809	2 F97183	membrane carboxype
52	104	6.9	901	2 F56833	phage infection pr
53	104	6.9	1474	2 T18281	hypothetical prote
54	104	6.9	2334	2 S32920	cell wall-associat
55	103.5	6.8	520	2 F70350	recombination prot
56	103.5	6.8	664	2 A37222	membrane associat
57	103.5	6.8	1964	2 A59282	nonmuscle myosin I
58	103.5	6.8	2017	1 A36014	myosin heavy chain
59	103.5	6.8	2057	2 S61477	myosin II heavy ch
60	103.5	6.8	2155	2 A02742	conserved hypotet
61	103.5	6.8	2155	2 C97523	hypothetical prote
62	103	6.8	631	2 JC4298	hyaluronan recepto
63	103	6.8	756	2 C64236	protein V (fcv) h
64	103	6.8	796	2 B84800	probable alpha-car
65	103	6.8	978	2 A70387	conserved hypotet
66	103	6.8	1070	2 F90106	IAP100 protein lim
67	103	6.8	2187	2 JC5837	364K Golgi complex
68	102.5	6.8	1961	1 A61231	myosin heavy chain
69	102.5	6.8	2022	2 T43214	ovtl protein - nem
70	102	6.7	451	2 T41722	probable gamma-glu
71	102	6.7	520	2 G71647	hypothetical prote
72	102	6.7	779	2 E97778	endopeptidase La (
73	102	6.7	1093	2 AC1753	tail protein [baet
74	102	6.7	1517	2 T13329	hypothetical prote
75	102	6.7	1976	2 A59252	myosin heavy chain
76	101.5	6.7	390	2 C90288	hypothetical prote
77	101.5	6.7	643	1 KERU2	keratin 1, type II
78	101.5	6.7	792	2 T49989	hypothetical prote
79	101.5	6.7	871	2 D86355	protein T16B15.12
80	101.5	6.7	982	2 I64232	protein p115 homol
81	101	6.7	1404	2 E36788	hypothetical prote
82	101	6.7	1938	2 I49484	alpha cardiac myos
83	100.5	6.6	320	2 B97206	methyl-accepting c
84	100.5	6.6	457	2 S2206	mesy protein - leu
85	100.5	6.6	550	2 B91286	probable membrane
86	100.5	6.6	550	2 F86137	hypothetical prote
87	100.5	6.6	1133	2 T22976	hypothetical prote
88	100.5	6.6	1147	2 T40866	cell polarity prot
89	100.5	6.6	1475	2 T33318	hypothetical prote
90	100	6.6	1046	2 A86790	ATP-dependent dADN
91	100	6.6	1188	2 G83960	chromosome segrega
92	100	6.6	1278	2 T27925	hypothetical prote
93	100	6.6	1701	2 A26868	major merozoite su
94	100	6.6	1738	2 T14857	interactin - slime
95	100	6.6	2677	2 D38194	desmoplakin 1 - hu
96	99.5	6.6	861	2 D82814	ATP-dependent Clp
97	99.5	6.6	1959	1 A33977	myosin heavy chain
98	99.5	6.6	2829	2 A42771	reticulocyte-bandi
99	99	6.5	772	2 A41860	tetracycline resis
100	99	6.5	971	2 A70179	exodeoxyribonuclea

ALIGNMENTS

RESULT 1

AE0673
hemolysin HlyE [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain C64864)
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AE0673
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mouton, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0673
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01758.1; PID:g16502606; GSPDB:GN00176
C:Genetics:
A:Gene: STY1498
C:Superfamily: Escherichia coli hemolysin E

Query Match 99.2%; Score 1503; DB 2; Length 305;
Best Local Similarity 99.7%; Pred. No. 4.2e-91;
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTSIFAEQTVVWVKSATETADGALDLYNKYLDQVVPKTFDETIKELSRFKQEQYSQASV 60
DB 3 MTGIFAEQTVVWVKSATETADGALDLYNKYLDQVVPKTFDETIKELSRFKQEQYSQASV 62
QY 61 LVGDIKLLMDSQDKYFEATQTVVWCGVATQLLAAVILLFDEYNEKKSAAQKIDILIRIL 120
DB 63 LVGDIKLLMDSQDKYFEATQTVVWCGVATQLLAAVILLFDEYNEKKSAAQKIDILIRIL 122
QY 121 DDGVKLLNEAQSLLTSSQSFNNASGKLLALDLSQLTNDFSEKSYFQSQVDRIKKEAYAG 180
DB 123 DDGVKLLNEAQSLLTSSQSFNNASGKLLALDLSQLTNDFSEKSYFQSQVDRIKKEAYAG 182
QY 181 AAAGIVAGPGLIISYSIAGVIEGKLIPELNRLKTVQNFFTLSATVKQANKDIDAAK 240
DB 183 AAAGIVAGPGLIISYSIAGVIEGKLIPELNRLKTVQNFFTLSATVKQANKDIDAAK 242
QY 241 LKLAETAAIGEIKETETETTFYVDYDMLSLKGAAXKMWINTCNEYQORHGKKTILFEV 300
DB 243 LKLAETAAIGEIKETETETTFYVDYDMLSLKGAAXKMWINTCNEYQORHGKKTILFEV 302
QY 301 PDV 303
DB 303 PDV 305

RESULT 2

C64864
hemolysin E - Escherichia coli (strain K-12)
N:Alternate names: hemolysin-inducing protein
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: C64864
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64864
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-305 <BLAT>
A:Cross-references: GB:AE000216; GB:U00096; NID:g1787417; PIDN:AACT4266.1; PID:g1787430;
A:Experimental source: strain K-12, substrain MGI655

C:Genetics:

A:Gene: hlyE; hpr

C:Function:

A:Description: hemolytic activity

A:Note: pore formation

C:Superfamily: Escherichia coli hemolysin E

C:Keywords: cytolysis; cytotoxin; hemolysis; transmembrane protein

F:181-197/Domain: transmembrane #status predicted <TM>

F:123/Active site: Asp #status predicted

Query Match 92.5%; Score 1402; DB 2; Length 305;
Best Local Similarity 91.1%; Pred. No. 1.7e-84;
Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

QY 1 MTSIFAEQTVVWVKSATETADGALDLYNKYLDQVVPKTFDETIKELSRFKQEQYSQASV 60
DB 3 MTEIVADKTVEWVKNATETADGALDLYNKYLDQVVPKTFDETIKELSRFKQEQYSQASV 62
QY 61 LVGDIKLLMDSQDKYFEATQTVVWCGVATQLLAAVILLFDEYNEKKSAAQKIDILIRIL 120
DB 63 LVGDIKLLMDSQDKYFEATQTVVWCGVATQLLAAVILLFDEYNEKKSAAQKIDILIRIL 122
QY 121 DDGVKLLNEAQSLLTSSQSFNNASGKLLALDLSQLTNDFSEKSYFQSQVDRIKKEAYAG 180
DB 123 DDGVKLLNEAQSLLTSSQSFNNASGKLLALDLSQLTNDFSEKSYFQSQVDRIKKEAYAG 182
QY 181 AAAGIVAGPGLIISYSIAGVIEGKLIPELNRLKTVQNFFTLSATVKQANKDIDAAK 240
DB 183 AAAGIVAGPGLIISYSIAGVIEGKLIPELNRLKTVQNFFTLSATVKQANKDIDAAK 242
QY 241 LKLAETAAIGEIKETETETTFYVDYDMLSLKGAAXKMWINTCNEYQORHGKKTILFEV 300
DB 243 LKLAETAAIGEIKETETETTFYVDYDMLSLKGAAXKMWINTCNEYQORHGKKTILFEV 302
QY 301 PDV 303
DB 303 PEV 305

RESULT 3

E30838
hemolysin E [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: E30838
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference strains
A:Reference number: A39629; MUID:21158231; PMID:111258796
A:Accession: E30838
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <HAY>
A:Cross-references: GB:BAC00007; PIDN:BAB35100.1; PID:g13361141; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: Ecs1677
C:Superfamily: Escherichia coli hemolysin E

Query Match 91.6%; Score 1387; DB 2; Length 305;
Best Local Similarity 90.1%; Pred. No. 1.6e-83;
Matches 273; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFAEQTVVWVKSATETADGALDLYNKYLDQVVPKTFDETIKELSRFKQEQYSQASV 60
DB 3 MTEIVADKTVEWVKNATETADGALDLYNKYLDQVVPKTFDETIKELSRFKQEQYSQASV 62
QY 61 LVGDIKLLMDSQDKYFEATQTVVWCGVATQLLAAVILLFDEYNEKKSAAQKIDILIRIL 120
DB 63 LVGDIKLLMDSQDKYFEATQTVVWCGVATQLLAAVILLFDEYNEKKSAAQKIDILIRIL 122
QY 121 DDGVKLLNEAQSLLTSSQSFNNASGKLLALDLSQLTNDFSEKSYFQSQVDRIKKEAYAG 180

```

Db 123 DGGITKLNDAKSLVSSQSFNNAGSKLLALDSQTNDFSEKSSYFQSQVDKIRKEAYAG 182
QY 181 AAAGVAGPFGIIISYIAAGVIEGKLIPELNNRLKTVQNFTLSATVKQANKDIDAAK 240
Db 183 AAAGVAGPFGIIISYIAAGVIEGKLIPELNNRLKTVQNFTLSATVKQANKDIDAAK 242
QY 241 LKLTETIAAIGIKETETETTRFVDDVDDMLSLKGAAKMINTCNEYQKRGKKTLPFV 300
Db 243 LKLTETIAAIGIKETETETTRFVDDVDDMLSLKGAAKMINTCNEYQKRGKKTLPFV 302
QY 301 PDV 303
Db 303 PEV 305

RESULT 4
E85696
probable pore forming hemolysin hlyE [imported] - Escherichia coli (strain O157:H7, subs
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85696
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Groetbeck, E.J.; Davis, N.W.; Lim, A.; Dimantanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85696
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <STO>
A:Cross-references: GB:AB005174; NID:G12514879; PIDN:AG56033.1; GSPDB:GN00145; UWGP:Z19
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: hlyE
C:Superfamily: Escherichia coli hemolysin E

Query Match 91.6%; Score 1387; DB 2; Length 305;
Best Local Similarity 90.1%; Pred. No. 1.6e-83;
Matches 273; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTSIFAEQTVVVVKSAIETADGALDLYNKYLQVTPWKTFTDIKELSRFKQESQASV 60
Db 3 MTEIVADKTVVVVKSAIETADGALDLYNKYLQVTPWKTFTDIKELSRFKQESQASV 62
QY 61 LVGDIKVLMDSDQKYFEATQTVYWCQVTVQVLLSAYILLDFDEYNEKASAKOILIRIL 120
Db 63 LVGNIKTLMDSDQKYFEATQTVYWCQVTVQVLLSAYILLDFDEYNEKASAKOILIKVL 122
QY 121 DGVKKLINEAQKSLTSSQSFNNAGSKLLALDSQTNDFSEKSSYFQSQVDKIRKEAYAG 180
Db 123 DGGITKLNDAKSLVSSQSFNNAGSKLLALDSQTNDFSEKSSYFQSQVDKIRKEAYAG 182
QY 181 AAAGVAGPFGIIISYIAAGVIEGKLIPELNNRLKTVQNFTLSATVKQANKDIDAAK 240
Db 183 AAAGVAGPFGIIISYIAAGVIEGKLIPELNNRLKTVQNFTLSATVKQANKDIDAAK 242
QY 241 LKLTETIAAIGIKETETETTRFVDDVDDMLSLKGAAKMINTCNEYQKRGKKTLPFV 300
Db 243 LKLTETIAAIGIKETETETTRFVDDVDDMLSLKGAAKMINTCNEYQKRGKKTLPFV 302
QY 301 PDV 303
Db 303 PEV 305

RESULT 5
T05634
hypothetical protein F20D10.190 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05634
R:Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.H
```

```

submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15420
A:Accession: T05634
A:Molecule type: DNA
A:Residues: 1-1496 <BEV>
A:Cross-references: EMBL:AL035538
A:Experimental source: cultivar Columbia; BAC clone F20D10
C:Genetics:
A:Map position: 4
A:Introns: 1042/2; 1212/2; 1232/1; 1263/3; 1349/3
A:Note: F20D10.190

Query Match 7.8%; Score 118; DB 2; Length 1496;
Best Local Similarity 19.6%; Pred. No. 10;
Matches 64; Conservative 67; Mismatches 122; Indels 74; Gaps 13;

QY 4 IFAEQTVVVVKSAIETADGALDLYNK--YLDQVTPWKT-----FDTIKELSRF-----KO 52
Db 153 IEAEKTVKGMKMRGRDDVVVKMEEEKSQVEEKLWKKEQFKHLEAEVEKLNLFKDSKK 212
QY 53 EYSQESVLVGDIKVL--LMDSDQKYFEATQTVYWC--GVVTQLLSAYILLDFEYNEKKA 109
Db 213 EWEESKSLDEIYSLQTKLDSVTRISEDLQKLLQMCNGALTQ-----EETRKH 263
QY 110 SAQKDLIRILDGVKKLINEAQKSLTSSQSFNNAGSKLLALDSQTNDFSEKSSYFQ-- 167
Db 264 EIQVSEFKAKYEDAFACQDARTQL-----DDLAKRDWEVAELRQTLNKKDAYFKEM 316
QY 168 ----SQVDRIRKEAYAGAAAGIVAGPFGIIISYIAAGVIEG---KLIPELNNRLKTVQN 220
Db 317 KYNGKLEQENRELLSLK-----ELQETIQQSGNSALSCLKNKKFRLNEN 362
QY 221 FTTLSLAT-----VKQANKDIDAAKLATEIAAIGIKETET-----TTRFV 264
Db 363 IHKNCANLRSKEAESSQVEKMWEEINDYKLOLQSKAALKVELELENCRSSTAKMRL 422
QY 265 DYDDLMLSLKGAAKMINTCNEYQOR 291
Db 423 QYBEISIMFL-----VLSRTVSEAQR 444

RESULT 6
LEBCA
hemolysin A - Escherichia coli
C:Species: Escherichia coli
C:Date: 30-Sep-1998 #sequence_revision 30-Sep-1998 #text_change 18-Jun-1999
C:Accession: A24433; I41280
R:Palmlie, T.; Pellett, S.; Welch, R.A.
J. Bacteriol. 163, 94-105, 1985
A:Title: Nucleotide sequence of an Escherichia coli chromosomal hemolysin.
A:Reference number: A24433; MUID:85234404; PMID:3891743
A:Accession: A24433
A:Molecule type: DNA
A:Residues: 1-1023 <REL>
A:Cross-references: GB:M10133; GB:M12863; NID:G146377; PIDN:AAA23975.1; PID:G146379
A:Experimental source: strain J96, O4 serotype
R:Stanley, P.; Packman, L.C.; Koronakis, V.; Hughes, C.
Science 266, 1992-1996, 1994
A:Title: Fatty acylation of two internal lysine residues required for the toxic activity
A:Reference number: A55387; MUID:95093925; PMID:7801126
A:Contents: annotation; lysine palmitoylation
A:Note: lysine modification is performed by the hlyC gene product
R:Haertlein, M.; Schiessl, S.; Wagner, W.; Rdest, U.; Kreft, J.; Goebel, W.
J. Cell Biol. 22, 87-97, 1983
A:Title: Transport of hemolysin by Escherichia coli.
A:Reference number: I41280
A:Accession: I41280
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1,'T',3,'V',5,'T',7-44 <RES>
A:Cross-references: GB:M29173; NID:G146337; PIDN:AAA23957.1; PID:G146338
C:Genetics:
A:Gene: hlyA
```


R.Kunst, P.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chedoke, A.A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati, E. *Nature* 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Y. M.; Ogawa, K.; Ogiwara, K.; Oudega, B.; Park, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, R.; Mew, M.; Mwa, R.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.M.; Portetelle, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Segkiguchi, J.; Sekowska, A.; Seron, A.; Authors: Schleich, S.; Schroeter, R.; Tanaka, T.; Terpsstra, P.; Tononi, A.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamanoto, H.; Yaman, K.; Yasumoto, K.; Yata, K.; Yoshida, K. *Artile*: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: F69913

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-478 <KUN>

A:Cross-references: GB:299115; GB:A009126; NID:G2634478; PIDN:CAE14031.1; PTD:el183560;

A:Experimental source: strain 168

C:Genetics:

Gene: yond

Query Match	7.6%;	Score	114.5;	DB	2;	Length	478;
Best Local Similarity	22.6%;	Pred.	No. 3.8;				
Matches	57;	Conservative	41;	Mismatches	73;	Indels	81;
Gaps	12;						
QY	25	DLYNKY-LDOVPWKTFDETIKELSPFKCEYSQEASVLVGDIKVLLIMDSODKYFEATQTIV	83				
Db	268	DYVDTVFVNVSWS--DENSYD-KYFKENYR-----TGDITSIDFSKTEVF-----WT	315				
QY	84	YEWCGVVTTOLLGSAIYLLDFEYNKKASAOKDLIRILDGGVKLNEAQSILTSSOSFPN	143				
Db	316	RWEEVVPPIOSQL-----NQKDSQIK--DLTKQVQNQINDKVIGIEQQFTNT	360				
QY	144	ASGKLIALDSQ-----LTNDFFSEKSYFSQVDRIKZAYAGAAAGIVAG	188				
Db	361	ASEKLVQLNSEVEQLPKYPEKEHKETLLQCKLEKNEFYKAKPEALNAEE-----	409				
QY	189	PGLLIISYIAAGVIEGKLIPELNNELKTVOFFTSLSATVQKANKDIAAKLKLATETA	248				
Db	410	-----KFS-----TEEVQNL---IHASVQK-DEEGEKAVLQNTMLDV	442				
QY	249	AIGEIXTETETT	260				
Db	443	DLVSVPTETNT	454				

RESULT 10
B59102
hypothetical protein pXO1-90 - Bacillus anthracis virulence plasmid pXO1
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: B59102
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXO1, the large Bacillus anthracis plasmid harbored
A:Reference number: A59091; MUID:99445493; PMID:10515943
A:Accession: B59102
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-652 xOKI>
A:Cross-references: GB:AF065404; NID:G4894216; PID:G4894306
A:Experimental source: strain Sterne
A:Note: similar to hypothetical protein; PFS0765W (980 aa); Plasmodium falciparum (AE001
C:Genetics:
A:Gene: pXO1-90
A:Genome: plasmid

Query Match	7.6%	Score 114.5;	DB 2;	Length 652;
Best Local Similarity	18.6%	Pred. No. 5.8;		

Matches	58;	Conservative	66;	Mismatches	106;	Indels	81;	Gaps	10;
QY	25	DLYNKYLQVLPWKTFDETIKELSRFKOEYSQEASVL-----VGDIKVLLMDSQDKYFE	78	:	:	:	:	:	:
Db	237	DVINQKIDB---FDKLNSCRKDLEMLMEELNGLSQLKKQSFLQLDLKNLKESESRLLE	293	:	:	:	:	:	:
QY	79	ATQ-----TYEWCGVVTVQLLSAYILLFDEYNE--KRASAQKDTILRIILDDGVKK	126	:	:	:	:	:	:
Db	294	LNNKDSNLELNSEIKLNDNRKAELSILIMELIKOOSFDFKKIKNEKDDLNNKRGEDLINR	353	:	:	:	:	:	:
QY	127	LNEAAQK-----SLITTSQQSFNNASGKILLALDSLQTLTNDFSEKSSYFQ-----SQV	170	:	:	:	:	:	:
Db	354	TAESKELAKKKAELNTKLVELFPVKQEALNKGSGQVLYIYNKLDNELRELADKYKNSDNKI	413	:	:	:	:	:	:
QY	171	DRIRKEAVAGAAAGIVAGPFGLIISYSTAAGVIEGKLIPELNNRLTKTVONFTTSLSATVK	230	:	:	:	:	:	:
Db	414	SRLKNH-----:::-----IGEYNKQLEKIEN-----ELE	435	:	:	:	:	:	:
QY	231	QANKDIDAAKLKLAT-EIAAIGEIKTETTTFYVDYDDL-----WLSLLKGAACKMWIT	284	:	:	:	:	:	:
Db	436	ECKNKIDNTKKQLAEFDSKNKKQOELESELVOLNKIDELGKHGHKHOEASQKKAJDE	495	:	:	:	:	:	:
QY	285	CNEYQORHGKK	295	:	:	:	:	:	:
Db	496	AKEINKKLAEK	506	:	:	:	:	:	:
 RESULT 11 AC1814									
hypothetical protein all0059 [imported] - Nostoc sp. (strain PCC 7120)									
CtSpecies: Nostoc sp. PCC 7120									
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120									
CtDate: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002									
C:Accession: AC1814									
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; et al.									
DNA Res. 8, 205-213, 2001									
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Nostoc sp. PCC 7120									
A:Reference number: AB1807; UID:21595285; PMID:11759840									
A:Accession: AC1814									
A>Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-727 <CUR>									
A:Cross-references: GB:BA000019; PID:GB17135037; GSPDB:GN00179									
A:Experimental source: strain PCC 7120									
C:Genetics:									
A:Gene: all0059									

Query Match	7.5%;	Score 113.5;	DB 2;	Length 727;
Best Local Similarity	22.2%;	Pred. No. 7.7;		
Matches	75;	Conservative 58;	Mismatches 134;	Indels 71; Gaps 15;
QY	5	FAEQTEVVVKSIAETADGALDLYNKYLDQVIEPWKTFDEIKELSRFKQBYSQASVVLVGD	64	
DB	194	FINARLPVIKQQLRQAEKLEQFRKKHLLDPELOSQVLIKSLTK-TQEQQLTRTQLQD	252	
QY	65	IKVLLMDSQDYFEATOIVYEGCVVQTLL---SAYILLFDEWEXKASQAQDILRIILD	121	
DB	253	INSRVQNUEERTEASQK-----ALVSMKLAOSSRYKTLTSELQKTEQSLAKEQL-RYTD	306	
QY	122	DG--VKKLINEAQKSILT-SSQSFNNASG-----KLIALDSQLTNDP-----	159	
DB	307	DSPIVQSLKQRRSQSLTVRQELKELTAEIKTQTTPEPQLGVGDPNLVEEFVQVCTTAL	366	
QY	160	----SEKSSYFQSDVRIRKEAVAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRL	215	
DB	367	GLIANEKS--LRSEQRIRSE-----LSKYPSLTAEYQ-----RLLPVETQR	407	
QY	216	KTVQGNFF-TLSATVKQANKQIDAKLKIATBIAIGELKTTETTRFVYVDDL-----	269	
DB	408	KTEQLQMQQQSMGMKIAHGSPD----WLVLBEANLG---TVVGNDRFVLVFGVMTGPI	460	
QY	270	---MLSLKGAAGKMINTCNEYQQRBHGKKTILEVPDVA	304	

Db 1940 EDQIAYSRNIEKARSBELNMNFKKINNTLAELPDLVEQCNITLL---YSQIDIEY 1996
Qy 105 NEK-----KASAKDILIRILD-----DGVKLNBAQKS 133
Db 1997 DEEVQTAGRAHAEKLEVOAQK-IVDRFVDRTRTENPLKASHAYENIVEALKNATEAVDS 2055
Qy 134 LLTSQSFNNASGKLLALDSQLTNDPSEKSYQSQVDRIRKAYAGAAAGIVAGPGLI 193
Db 2056 AAEASE-----AVSKWLGSEGESDANEES--LRSQLEKLNKNESSJN-----2097
Qy 194 ISYGIAGVIE--CKLIPELNNRLKTVQNFPTLSATVQKQKIDDAKCLKLATEIAAIG 251
Db 2098 VDNNAVKIVEELKKEKKDLTRLGHLNELKTSI-----VKRLG 2136
Qy 252 EIKTETETTRPYVDYDMLSLKGAAKM-INTCNEYQORHGKKT 296
Db 2137 VIKNEASS--WDDKDRMHSILKNGAKTAHARSANVKESEGIKT 2179
RESULT 15
T43291
laminin alpha chain - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T43291
R:Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang
submitted to the EMBL Data Library, June 1998
A:Description: Expression, function and evolution of laminin alpha chains.
A:Reference number: 422397
A:Accession: T43291
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3102 <ZHU>
A:Cross-references: EMBL:AF074902; PIDN:AAC26793.1
C:Genetics:
A:Map position: 1
A:Note: lamal/2
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h
Query Match 7.5%; Score 113.5; DB 2; Length 3102;
Best Local Similarity 20.8%; Pred. No. 51;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;
Qy 25 DLYNKYLDQVTPW-KTDFETIKELSRFKQEVQSQASVLVG-----DKVLLMDS 72
Db 1881 DLKNR-IDVLEQWMDYRTIYDVSKKDTADAEMSLVVGKRNRYKEVSEIEKLVEA 1939
Qy 73 QDK-----YFEATQTV-----YEWQGVVTQLLSAVILLFDEY 104
Db 1940 EDQIAYSRNIEKARSBELNMNFKKINNTLAELPDLVEQCNITLL---YSQIDIEY 1996
Qy 105 NEK-----KASAKDILIRILD-----DGVKLNBAQKS 133
Db 1997 DEEVQTAGRAHAEKLEVOAQK-IVDRFVDRTRTENPLKASHAYENIVEALKNATEAVDS 2055
Qy 134 LLTSQSFNNASGKLLALDSQLTNDPSEKSYQSQVDRIRKAYAGAAAGIVAGPGLI 193
Db 2056 AAEASE-----AVSKWLGSEGESDANEES--LRSQLEKLNKNESSJN-----2097
Qy 194 ISYGIAGVIE--CKLIPELNNRLKTVQNFPTLSATVQKQKIDDAKCLKLATEIAAIG 251
Db 2098 VDNNAVKIVEELKKEKKDLTRLGHLNELKTSI-----VKRLG 2136
Qy 252 EIKTETETTRPYVDYDMLSLKGAAKM-INTCNEYQORHGKKT 296
Db 2137 VIKNEASS--WDDKDRMHSILKNGAKTAHARSANVKESEGIKT 2179
RESULT 16
S30834
hypothetical protein YEL043w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae

C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 19-Apr-2002
C:Accession: S30834; S50501
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor,
submitted to the EMBL Data Library, February 1993
A:Reference number: S30812
A:Accession: S30834
A:Molecule type: DNA
A:Residues: 1-956 <MUL>
A:Cross-references: GB:U18779; EMBL:L10830; NID:G603625; PIDN:AAB64999.1; PID:G603636
R:Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.
A:Reference number: S50491
A:Accession: S50501
A:Molecule type: DNA
A:Residues: 1-956 <DIE>
A:Cross-references: EMBL:U18779; NID:G603625; PIDN:AAB64999.1; PID:G603636; MIPS:YEL043w
C:Genetics:
A:Cross-references: SGD:S0000769
A:Map position: 5L
Query Match 7.5%; Score 113; DB 2; Length 956;
Best Local Similarity 24.4%; Pred. No. 12;
Matches 68; Conservative 45; Mismatches 78; Indels 88; Gaps 15;
Qy 38 KTFDETIKELSRFKQEVQSQASVLVGDKVLLMD-----SODKYFEATQTVYEWQGVVTO 92
Db 272 KSLKSNISL-----ENSKLTDLKIENKNDKSKKIKSVKQNDQKKSQEDTE 322
Qy 93 LLS-----AVILLFDEYNKKASAKQDI-----LIRLDGVKKKNEAQSLLTS---S 138
Db 323 LLSKDTIKETKFKLINESNASVINKEIESLQNEISKWEESKRLNASKSLITSIVN 382
Qy 139 OSFNN-----ASGKLALDSQLTNDFS-EKSYF-----QSQVDRIRKAYA 179
Db 383 ANVENDKPIASGELSAVLKLL-NDFTLEKNGFLNAGEEFLSKLNADSSLLFMKQE--- 438
Qy 180 GAAAGIVAGPGLIISYIAA-----GVIEGKLIP-ELNNR-LKT-----V 218
Db 439 -----LSIDQLEANWKLQSRNLLKISALENQFNEMSLNNRLKTKLMVQPY 486
Qy 219 QNFFTLSLAT-----VKQANKDIDAAKLKLAIEAAGEI 253
Db 487 KNGDSLAAATNNSNAEKNSRSGSIQPLSNMBSRTGSI 525
RESULT 17
A64465
hypothetical protein MJ1322 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: A64465
R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
isen, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaize, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: A64465
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1005 <BUL>
A:Cross-references: GB:U67572; GB:L77117; NID:G1591958; PIDN:AAB99331.1; PID:G1591962; T
C:Genetics:
A:Map position: REV1273394-1270377
C:Superfamily: hypothetical protein MJ1322
Query Match 7.5%; Score 113; DB 2; Length 1005;
Best Local Similarity 19.7%; Pred. No. 13;
Matches 60; Conservative 55; Mismatches 91; Indels 98; Gaps 12;
Qy 7 EGTVEVWKSATETADGALDLYNKYLDQVLPWKTFDE-----TIKELSRF-----KQEYSQ 56

Db 605 DEILEDIKSQLNKEK---NFYQYLSAVSYLVNSVDREGIRNRKEIENIVSGWKEKCRE 661
Qy 57 EASVY-----VGDIVLLMSQDKYFEATQTYEWGVVTVQLLSAYILLFDEYNKKAS 110
Db 662 ELNKLREDEIRNRLKDKLNELNKEKE-----LIETENRRSLKFKYKEYLGL 710
Qy 111 AQDILLRILDDGVKKLNEAKSKLTSQSFFNNSAGKLLALD----- 152
Db 711 TEKLEEKXNKDGLLEI-----YINCNSKILAIIDNKKYKEDIEIYLNKK 757
Qy 153 ----SQTNDPSEKSSYFQSVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLI 208
Db 758 ILEVKEINDEIRISYINQKLE-----INYNEE-----EHKKI 792
Qy 209 PEL--NNR--LKTQNFPTSLSATVQKANKDIDAAXKLKLAETAAIGEIKTETETTRFYV 264
Db 793 KELYNKQRLNDVREOKTETIETGIEYLLKDVESLKLRL-----KEMSNLEKEKLUKTRV 848
Qy 265 DYDD 268
Db 849 EYLD 852

RESULT 18

hemolysin A - Escherichia coli plasmid phly152
C:Species: Escherichia coli
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 04-Mar-2000
C:Accession: S10056
R:Hees, J.; Wells, W.; Vogel, M.; Goebel, W.
FEMS Microbiol. Lett. 34, 1-11, 1986
A:Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison with the hemolysin A gene of Escherichia coli
A:Reference number: S07209
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1024 <HES>
A:Cross-references: EMBL:M14107
C:Genetics:
C:Genome: plasmid phly152
C:Superfamily: hemolysin A; hemolysin A homology
C:Keywords: lipoprotein
F:247-792/Domain: hemolysin A homology <HLVA>
F:564,690/Binding site: palmitate (Lys) (covalent) #status predicted

Query Match 7.4%; Score 112.5; DB 2; Length 1024;
Best Local Similarity 20.8%; Pred. No. 14;
Matches 70; Conservative 55; Mismatches 115; Indels 97; Gaps 11;
Qy 2 TSIPAEQTVVVK-----SAIETADGALDLYNKYLDQVTPWKTFDETIKELSRFKOE 53
Db 155 TALSSMKIDELIKKQSGNVSSSELAKASIELINQLVDIVASIN-----NNVNSFSQ 208
Qy 54 YSQEASV-----VGDIVLLMSQDKYFEATQTYEWGVVTVQLLSAYILLFDEY 104
Db 209 LNTGSLVSLTKHLNGVGN-KLQPLNDINIGAGLTV-----SGILSAISAFILSNADAD 264
Qy 105 NEKASAKQDILRIILD---GVKKLNEAKSKLTSQSFFNNSAGKLLALDLSQLTNDPSE 161
Db 265 TRTKAAGVELTKVLGVNGKISQYIIAQRAGCLSTS----- 303
Qy 162 KSSYFQSVDRIRKEAYAGAAAGIVAGPGLIIS-----YSIAAGVIEGKLIPELNNRLKT 217
Db 304 -----AAAAGLIASAVTLAISPLSIFLSIAADKFRANKIEYSORFKK 345
Qy 218 VONFPTSLSATVQKANKDIDAAXKLKLAETAAIGEIKTETETTRFYVDVLLMLSLKGA 277
Db 346 LGYDGSLLAFHKTGAIDASLTITSTVLASVSSGISAATISLV-----GAPVSALVGA 401
Qy 278 -----AKQMTNCNEYQQRHGK 294
Db 402 VTGIISGILEASKQAMFEHVASKVADVIAEWKKGK 438

RESULT 19

S18199
myosin heavy chain - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 02-Feb-2001
C:Accession: S18199
R:Stewart, A.F.R.; Camoretti-Mercado, B.; Perlman, D.; Gupta, M.; Jakovcic, S.; Zak, R.
J. Mol. Evol. 33, 357-366, 1991
A:Title: Structural and phylogenetic analysis of the chicken ventricular myosin heavy chain
A:Reference number: S18199; MUID:92130260; PMID:1774788
A:Accession: S18199
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1039 <STE>
A:Cross-references: EMBL:X59552; NID:g62995; PIDN:CAA42130.1; PID:g62996
A>Note: In the authors' translation 45-Lys is shown after residue 40, and, consequently,
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: ATP; nucleotide binding; P-loop
F:402-409/Region: nucleotide-binding motif A (P-loop)

Query Match 7.4%; Score 112; DB 2; Length 1039;
Best Local Similarity 19.9%; Pred. No. 15;
Matches 69; Conservative 67; Mismatches 143; Indels 68; Gaps 11;

Qy 1 MTSIFAEQTVVVKSAIETADGALDLYNK--YLDQVTPWKT-FDETIKELSRFKQEYSQE 57
Db 369 MTRLMNDLTTQTKLOSGENGFEVQLBEKESLISQLSRGKTSFTQIIEELRRQLLEETKS 428
Qy 58 ASVLVGDIVKLLMD-----SODKYFEATQTYEWGVVTVQLLSAYILLFDEYNEK 107
Db 429 KNAIAHALQAARHDCDLLREYEEQAKAELOALSAGNAEVAQWTKVETDAIQRTTE 488
Qy 108 KASAQKQDILIRI-----LDDGVKKLNEAKSKLTSQSFFNNSAGKLLAL 151
Db 489 LEDAKKLLARLQEAEEAIEAANAKSSLEKAKHRLQNEQEDMMDLEKANSAAA---SL 545
Qy 152 D-----SOLNDFSEKSSYFQSVDRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEG 205
Db 546 DKQRGFDKIINDWKQKYEESQAELEASQKE-----ARSLSTELFKLKNAYEETLDHLE- 599
Qy 206 KLIPELNNRLTKVQNFFTSISATVQKANKDI-----DAAKLKLATEIAAIGE 253
Db 600 ----TLKEENKNLQEEISDLTNQISEGNLNLHETKVKQVEQEKSEVQLALEAEAGALE 655
Qy 254 KTEETTRFYVDVLLMLSLKGAAKKMTNCNEYQQ---RHGKKTL 297
Db 656 HEESKTLRF-----QLELSQLKADFERKLAEKDEEMQIRRNQORTI 697

RESULT 20

S21801
myosin heavy chain, neuronal [similarity] - rat
N:Alternate names: myosin II
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: S21801; PNO013; S18134
R:Sun, W.; Chantler, P.D.
J. Mol. Biol. 224, 1185-1193, 1992
A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain
A:Reference number: S21801; MUID:92235856; PMID:1569576
A:Accession: S21801
A:Molecule type: mRNA
A:Residues: 1-1999 <SUN>
A:Cross-references: EMBL:X62659
R:Sun, W.; Chantler, P.D.
Biochem. Biophys. Res. Commun. 175, 244-249, 1991
A:Title: A unique cellular myosin II exhibiting differential expression in the cerebral
A:Reference number: PNO013; MUID:91151356; PMID:1998509
A:Accession: PNO013
A:Molecule type: mRNA

hypothetical protein - Synchocystis sp. (strain PCC 6803)

C:Species: Synchocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S75986
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, N. Reg. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis sp.
S:Reference number: S74322; MUID: 97061201; PMID: 8905231
A:Accession: S75986
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-584 <KAN>
A:Cross-references: EMBL:D64006; GB:AB001339; NID:gl001291; PIDN:BAAL0833.1; PID:gl00134
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: arginine-trna ligase

Query Match 7.3%; Score 110.5; DB 2; Length 584;
Best Local Similarity 23.4%; Pred. No. 9.1;
Matches 73; Conservative 50; Mismatches 128; Indels 61; Gaps 15;
QY 11 EVVKSATETADGALDLYNKYLDQVTPKTFDETIKLSRPFQYSQEAFLV--GDIK-- 66
DB 179 EVTFEALVTAD-ALDIG-----DLVTFYKQAKQRFDDQFR-ETSRQAVVALQAGDAKSI 232
QY 67 --VLLMDSQDKFEATQTVYEWGCVVTVQ-----LLSAYILLFDEYNEKKAQKD 114
DB 233 KAWQLLCEQSRREF--QLIYDCLDITIEBGSFYNPLPGVVLLQE-----KD 280
QY 115 ILIR-----ILLDGVKLEAKSLLT--SSQSFNNAAGKLLALDQLTNPSEKSSYP 166
DB 281 LLVEDNAGQCVFLDGFNTKGDRLPLIVQKSDGYNATDLAALNRYLNTDGAEXIIV 340
QY 167 --QSQVDRIRKEAYAGAAAGVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFETS 224
DB 341 TDAGQANHEAQQFQVAEKAGILTDPTQV--HVPFGLVKGEDGKLLKTRAGDTIELKDL 397
QY 225 LSATVQKAKDID-----AKLKLATBIA--AIGEIK-----TETTRFVYVDVDDLM 270
DB 398 LTEAVTRAQDLETRLFAERSETEEFKEVQAVGVIGAVKADLSQNRSTSYVFSFKM 457
QY 271 LSLKLGAAKXMI 282
DB 458 LALQGTAPYML 469

RESULT 24
AG1739
transmembrane protein [imported] - Listeria innocua (strain Clp11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG1739
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.: Title: Comparative genomics of Listeria species
A:Reference number: AB1077; MUID: 21537279; PMID: 11679669
A:Accession: AG1739
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-927 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97687.1; PID:gl6414982; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin2460

Query Match 7.3%; Score 110.5; DB 2; Length 927;
Best Local Similarity 20.6%; Pred. No. 17;

Matches 79; Conservative 45; Mismatches 132; Indels 127; Gaps 16;
QY 1 MTSIPATQTEVVKSAIETADG---ALDLYNKYLDQVTPW---KTPDETIKLSRPFQ 53
DB 160 VTKSYAEAFIDKIK---ESGDGFAQAADGSGKIKDGLVKSQEGNKTTISTNLKTLADSSLT 216
QY 54 YSQEASVL-----VGDIK-----VLLMDS 72
DB 217 FYDQANTLEVGLKTYTQVNTAAAGGDKLNAGVSTLAAGVPLKDGVAALDGGATKLSGG 276
QY 73 QDKYFEATQTVYEWGCVVTVQLLSAYILLFDEYNEKKAQAQKDILIRLLDGVKLEAKQ 132
DB 277 VSTYTSQVDTL---SGGINQAYNGSTALSGLNKGNGS-----VPALASGVTVQLNDGQK 327
QY 133 SLLTSSQSFNNAAGKLLA---LDSQLTN-----DFSEKSSYFQSOVDRIKAEAYAGAA 182
DB 328 SLATGLNSLDVSGNKLKAGLKELDGNLTNSQKIAQLKQGNLDLQGGIDQLNQNVGEDA 387
QY 183 AGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQ-----NFTTSLSAT 226
DB 388 A-----LAKQLQA---LQKSLTDLQGLTTFIKSNANFADAEAIKAKINATNGVSAE 434
QY 229 VKQ-----ANKDIDAALKLATBIAAIGEIKT-----ETETTRFVYVDVDDMLSL 273
DB 435 DKQKLIIDAIQADLDKESQKSAQVATVEKLSQSGLSGLDLAAIQTVT-----EL 483
QY 274 LKGAAKMINTCNEYQQRHGKKT 296
DB 484 QTGVAKISAG---YKAVHGGTT 502

RESULT 25
S67087
hypothetical protein YOR195w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O4806
C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
C:Accession: S67087
R:Hughes, B.; Pohl, T.M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66685
A:Accession: S67087
A:Molecule type: DNA
A:Residues: 1-821 <HUG>
A:Cross-references: EMBL:Z75103; NID:gl420464; PID:e252389; PID:gl420465; GSPDB:GN00015;
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:SLK19; MIPS:YOR195w
A:Cross-references: SGD:S0005721
A:Map position: 15R

Query Match 7.3%; Score 110; DB 2; Length 821;
Best Local Similarity 20.4%; Pred. No. 15;
Matches 68; Conservative 52; Mismatches 109; Indels 104; Gaps 12;
QY 10 VEVVKSATETADGALDLYNKYLDQVTPKTFDETIKLSRPFQYSQEAFLVGDIKVL- 68
DB 373 VEKFKRIKELNTEIKVLNS--NOKLQEKFDASITEVNHKGEHENTVNTLQONKILN 430
QY 69 -----LMDSDQKYPEATQTVYEWGCVVTVQLLSAYILLFDEYNEKKAQAQKDIL 116
DB 431 DKNVELENKAKELKNDKLSYEYETTLNLSRIVQL-----NDKIET--DIV 477
QY 117 IRLDGVKLEAKSLLTSSQSFNNAAGKLLALDQLTNDPSEKSSYFQSOVDRIKAE 176
DB 478 LKSKENELDNKLKLSIKETLSIKDFNDS-----LIQINELISTKN-LQKND----- 526
QY 177 AYAGAAGIVAGPGLIISYSIAAGVIEGKLIPELNN-----RLKTVQNFETSLSATVKA 232
DB 527 -----DNNLNDNLKVVQDKLIKNEETLKLK 553
QY 233 NKDIDAAKL---KLATEIAA-----IGEIKTET-----ETTRFVYVDVDDL----- 269

A:Molecule type: DNA
A:Residues: 1-1545 <LIN>
A:Cross-references: EMBL:AF074611; NID:g3883003; PID:g3883049; PIDN:AAC82709.1
C:Genetics:
A:Gene: Y1049
A:Genome: plasmid pMT1

Query Match 7.2%; Score 109.5; DB 2; Length 1545;
Best Local Similarity 19.7%; Pred. No. 38;
Matches 57; Conservative 54; Mismatches 101; Indels 77; Gaps 9;
QY 2 TSFAETVVKSAIETADGALDLYNKLDQVTPKTFDETIKELSRFKQYSQASVL 61
DB 849 SSELGQELLEIDA--KASQDAVDATNKQMEEL--KELDQSVADLDSKLEDTSGLEQV 904
QY 62 VGDIKVLLMDSQDKYFEATQTVVWCGVVTQLL-----SAYIL 99
DB 905 QNDLKNEVSGTLKNDALQVDSNAALVELQETVSEQKATAGAVEAAHAALDNASAL 964
QY 100 LFDEYNEK----KASA-QKDILIRILDGKVKLNEAKSLLTSQSFNNASGKLIAL----- 151
DB 965 IASERARVEGDKANAKQIEAMKSSVDDSVAAVEEMKKTVAEYERASAEASTNIEALAKT 1024
QY 152 -----DSQLTND-----FSEKSYFOSQVDRIKKEAYAGAAAGIVAGPF 190
DB 1025 NIDLALRQDEDOHKQWNNAKIATTKTFADDSAMASKVEIRAE----- 1070
QY 191 GLIISYISGAVTEGKILPELNNRLKTVQNFTSLGATVKQANKIDAA 239
DB 1071 ---IGEDIRASILE-----ETARVEADKTIATHISKLEAQLNDDISAA 1111

RESULT 29
P75008
hypothetical protein PAB1294 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: P75008
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: P75008
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-595 <KAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50508.1; PID:g545902
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1294

Query Match 7.2%; Score 109; DB 2; Length 595;
Best Local Similarity 23.0%; Pred. No. 12;
Matches 71; Conservative 55; Mismatches 109; Indels 74; Gaps 18;
QY 24 LDLYNKLDQVTPKTFDETIKELSRFKQYSQASVLVG-DIKVLLMDSQDKYFEATQ 82
DB 45 LDLASESLDQIEGNSNSTSY-ELSRILNKTREAYIYKSRGIETKVLDPPLPKLSEG 103
QY 83 VYEWCGVVTQLLSAYILLFDEYNEK---ASAQKOILIRILDGKVKLNEAKSLLTS 138
DB 104 I---GNILKGRSSFLMYVSNFQXNEMPELITAMEESLI-IMDDGIK---EAKLALIRIS 155
QY 139 QSFNNASGKLIALD-SQLTNDPSEKSYFOSQVDRIKKEAYAGAAAGIV-----AGPFL 192
DB 156 KIEFSGAGKLDVNDILEKLNKLERESSYHKILERL-KSRVTPYRKLILFVSNAPYVL 214
QY 193 --IISYSIAAG-----VIEGKILPELNNRLKTVQNFTSL-----SATVKQANK 234
DB 215 ENVTYFYFKNASSVQILINGSILK-----ANVDNGFFSVNYSPFLPGIYSAAVASKGNL 268
QY 235 D-----IDAAKLK---LATET-AAIKEIKTETETTR-----FVVDVDDLMLSL 273

DB 269 TSNTVILNVSKRTVFPVVPSEVYGRIGETVTVISGFLRDLNLGYPGPKIEVIDY----- 321
QY 274 LKGAAXKXMI 282
DB 322 -KGRATKLI 329
RESULT 30
P75001
methyl-accepting chemotaxis protein (tlpc) PAB1336 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: P75001
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: P75001
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-739 <KAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50454.1; PID:el51635;
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1336

Query Match 7.2%; Score 109; DB 2; Length 739;
Best Local Similarity 19.2%; Pred. No. 16;
Matches 64; Conservative 71; Mismatches 138; Indels 60; Gaps 11;
QY 1 MTSIFAETVVKSA--IETADGALDLYNKLDQVTPKTFDETIKELSRFKQYSQEA 58
DB 425 VTEFRSIGSLVEMANDLEKRNALAQSKDVTEAL--NQVNEAIQOVSIERQOQETI 482
QY 59 SVLVGDIKVLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDI--- 115
DB 483 NEITDGRVLVAQTSE---ESVRAMEBFSGAVTEWS---IANEGSQKGDALKRIEDI 534
QY 116 -LIRILDGKVKLNEAKSL-----LTSQSNNASGKLALDS----- 153
DB 535 QHMSRIETVSKVAENSRNIEETNVTISAEQTN---LLALNAAIEARAGEAGRGF 590
QY 154 -----QLTNDPSEKSYFOSQVDRIK---KEAYAGAAAGI-VAGPFLIISYSTAAG 201
DB 591 AVVAQEIRKLAESKQAAADNIKSIDKITDIEKAEVATEKGVSVIGESSETLDTTIGYL 650
QY 202 VIEGKILPELNNRLKTVQNFTSLGATVKQANKIDAAKLKLAETIAAIGIKTETETTR 261
DB 651 ANIATLQETSERMTYKE-----QIVRTQESVDKALALENLAASAEETASAEVS 703
QY 262 FVVDVDDLMLSLKGAANKMINTCNEYQQRHGK 294
DB 704 SAIEQQTAAIEELRRAAQELKDMVGRMRQIVGK 736

RESULT 31
A3254
leukotoxin A - Pasteurella haemolytica (serotype T10)
N:Alternate names: lktA protein
C:Species: Pasteurella haemolytica
C:Date: 10-Aug-1990 #sequence_revision 15-Nov-1996 #text_change 18-Jun-1999
C:Accession: S37145; A35254; S34237; S34235
R:Lainson, A.F.; Aitchison, K.; Donachie, W.
submitted to the EMBL Data Library, September 1993
A:Description: DNA sequence of the leukotoxin A gene from P. haemolytica T10 serotype.
A:Reference number: S37145
A:Accession: S37145
A:Molecule type: DNA
A:Residues: 1-955 <LAI>
A:Cross-references: EMBL:226247; NID:g400424; PIDN:CAAB1206.1; PID:g400425
R:Highlander, S.K.; Engler, M.J.; Weinstock, G.M.
J. Bacteriol. 172, 2343-2350, 1990
A:Title: Secretion and expression of the Pasteurella haemolytica leukotoxin.

Query Match	7.1%;	Score 107.5;	DB 2;	Length 1875;
Best Local Similarity	19.9%;	Pred. No. 65;		
Matches	64;	Conservative	49;	Mismatches 127; Indels 81; Gaps 12;
QY	7	EQTVVVKSAIETADGALDLYNKVLQDVIPWKTFTDTI	-----KELSRFKQVYSQES	59
DB	966	EDKISLLKEQFNLNEDLDQKGMK- -EKADFKRISLONNNKEVEAVKSEYSKLS	1023	
QY	60	VLVGDIKVLMDSDQKYFEATQTVYECGVVTVLLSAYILLFDYNEKKASAKQDI	----	115
DB	1024	KIQND- -LDQOTIYANTQNNYE	-----QELQKHADVSKTISELR	1061
QY	116	LIRILDDGVKKLN- -EAKSILLTSSQSFNNASGKLLA- -LDSQLTNDFEKSYFQ	167	
DB	1062	EQHTYKGQVKTNLNSRDQJENALKENKESWSQKSLLEQLDLSRIEDLSQKLLY	1121	
QY	168	SQVDRIRKEAYAGAAGIVAGPFLGIISYIAAGVIEGKLIPELNNRLKTVQNFFTLSA	227	
DB	1122	DQI- -QIYTAADKEVNNSTNG- -	-----PGLNNILITRRERDILDT	1159
QY	228	TVKQANKDIDAAKLKLAETAAIGETETETTR- -	-----FVVDYDDLM- -LSLL	274
DB	1160	KVTVAERDAKMLKOKLSLMDVELODAKTLKNSRVEKENHSSIIQOHHIMEKLNQLNLL	1219	
QY	275	KGAACKWINTCNEYQQRHGKK	295	
DB	1220	R- -ESNITURLENNNNKK	1237	
RESULT 39				
F72288				
methyl-accepting chemotaxis protein - Thermotoga maritima (strain MSB8)				
C;Species: Thermotoga maritima				
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000				
C;Accession: F72288				
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.A.; C.M.				
Nature 399, 323-329, 1999				
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq				
A;Reference number: A72200; MUID:99287316; PMID:10360571				
A;Accession: F72288				
A;Status: preliminary				
A;Molecule type: DNA				
A;Residues: 1-539 <ARN>				
A;Cross-references: GB:AE001772; GB:AE000512; NID:g4981693; PIDN:AAD36222.1; PID:g4981693				
A;Experimental source: strain MSB8				
C;Genetics:				
A;Gene: TW1146				
C;Superfamily: methyl-accepting chemotaxis protein				
Query Match	7.0%;	Score 106.5;	DB 2;	Length 539;
Best Local Similarity	19.4%;	Pred. No. 15;		
Matches	59;	Conservative	63;	Mismatches 147; Indels 35; Gaps 10;
QY	10	VEVYKSAIETADGAL- -DLYNKYLQDVIPWKTFTDTIKELSRFKQVYSQESVLYGDIK	66	
DB	206	IEKIRKDEITGKAAMAVEKLRILLDDITGINKASSEVSSSSSELSATSELSANVNSIS	265	
QY	67	VLLMDSDQKYFEATQTVYECGVVTVLLSAYILLFDYNEKKASAKQDILIRLD	----	121
DB	266	EALVSLNKADENSATLEEFTASIEELST- -	-----ADSNKSAQAMLESTQVHEQVEKS	320
QY	122	DGVKLNLEAKQSLLTSSQSFNNASGKLLADSLT	-----NDFSEKSSY- -FQSOVDR	172
DB	321	TERIREITEKAHSTRESENTKQALNRLLSMAEWINSIVDTINSIAEQTNLLALNAIEA	380	
QY	173	IRKEAYAGAAGIVAGPFLGIISYIAA- -	-----GVIEGKLIPELNNRLKTVQNFFTSL- -S	226
DB	381	AR-AGENAGRGFAVVVAIEIRKLAIESKAAATQOIGELGKURDEINNSKIVESTASAEET	439	
QY	227	ATVQANKDIDAAKLKLAETAAIGETETETTRFVVDYDDLMLSLLKGAQKQVINTCN	286	

```

Db      440 ASLVESIKDV-FESIRAME-----DVQSRVESVAASTOQSASLEELSAGVTRLTLLN 493
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      287 EYQQ 290
      : :
Db      494 KTRF 497

RESULT 40
AH1369
transmembrane protein [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH1369
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Ameid, A.; Baquero, F.; Bezche, P.;
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entian, K.D.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournan,
ck, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.;
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1369
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-927 <GUA>
A:Cross-references: GB:NC_003210; PIN:CAD00436.1; PID:G16411849; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2350

```

RESULT 41

REF ID: A60593

hypothesis i

C; Species

C; Date: 2
C: Accessi

R: Chambaur

Nucleic A

A;Title:

A;Reference number: A99512; NUID:21267165; PMID:11353084
A;Accession: C90593
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1272 <KUR>
A;Cross-references: GB:AI445566; PID:gl14090066; PIDN:CACI
A;Experimental source: strain UAB CT1P
C;Genetics:
A;Gene: MYPU 6510
A;Genetic code: SGC3

```

Query Match      7.0%; Score 106.5; DB 2; Length 1272;
Best local Similarity 23.1%; Pred. No. 46;
Matches 52; Conservative 41; Mismatches 91; Indels 41; Gaps 9;

Qy      14 KSAIETADGADLVNKNYLQDVI PWKTPDETIKELSRKQEYS---QEASVLVAGDIKVLML 70
      438 KPKVEPKDETALIPDK-ISKI-----ELKENSOLKQKLPSQFKESDLSNLKVLVS 488

Qy      71 DSQKYFEATQTVYEWCGVVTQLSAYILLPDEYNKKASAQKQILIRILDGQVKKLNKA 130
      489 DDKNKFSELS-----LPQGYISPKLASNNDEGTLIDVKVIVQKGKVEKTKELK 539

Qy      131 QKSLLTSSQSFNNASGKLLADSLQTDNDFSEKSYVFQSOV--DRIRKEAYAGAAAGTVA 187
      540 LTNLLTFESLKES-----DFQJ--DFSNNKQRLASVVMNDKLIKESLVVKNKTIE 590

Qy      188 GPGGLI-ISYSIAAGVIEGKIULPELNNRLKTVQNFFTSLSATVQK 231

Db      591 PDFNKYIISYGVSS-----LDEVNGKLIKWTIFKTKDKRLKE 628

```

RESULT 42

A35186
salivary agglutinin receptor precursor - Streptococcus sanguis
C:Species: Streptococcus sanguis
C:Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 12-Apr-1995
C:Accession: A35186
E:Demuth, D.R.; Golub, E.E.; Malamud, D.
J. Biol. Chem. 265, 7120-7126, 1990
A:Title: Streptococcal-host interactions. Structural and functional analysis of a Streptococcal cross-reactive agglutinin.
A:Reference number: A35186; MUID:90236997; PMID:2185241
A:Accession: A35186
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1473 <DEM>
A:Cross-references: GB:J05418
C:Superfamily: surface antigen spap

Query Match	7.0%;	Score 106.5;	DB 2;	Length 1473;
Best Local Similarity	23.5%;	Pred. No. 56;		
Matches 78;	Conservative 42;	Mismatches 97;		
		Indels 115;	Gaps 18;	

QY 15 SA IETADGALDLYNKYLDQVIPWKTFDETIKELSRFKQE---YSQEASVLVGDIKVLMD 71

Db 123 TATDNAQKQDEIKSDYAKQAEI KTTTEA-----YKKEVAHQAE TDKINAENKA----- 172

QY 72 SQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASQKDILIRILDGDKLNEAQ 131

173 ADDKYQDKLSHQ3--EVEKINTANATAKAEEYAKLAQYQKDLAT-----VKKANE-- 221

QY 132 KSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQVDRI-----KEAYAGAAAGI 185

Db 222 -----DSQ--QDYQNKL SAYQT ELARVQKANA EAK EYKAVKEN 259

QY 186 VAGPFGLIISYSIAAGVIEGKLIPELNNRLK-----TVQNFFTSLSATVKQANKDIDA-A 239

Db 260 TAKNEAL-----KVENEAIKQNETAKATYEAAMKQYEADLAA-IKKANEDNDADY 309

Qy 240 KLKLA--TEIAAIGE-----IKTETETTRFVD-- 265

Db 310 QAKLAAYQTELARVQKANAEEAYDKAVKENTAKNTAIQAENEAIKQRNETAKATYDAA 369

Qy 266 ---YD-DLMLSLKGAQKQWINTCNE--YQQR 291
|: || || || || || || || ||
Db 370 VKKYEADL-----AAVKQANATWEADYQAK 394

RESULT 43

Tl13829
Tpr homolog - fruit fly (*Drosophila melanogaster*)
C/Species: *Drosophila melanogaster*
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C/Accession: Tl13829
R:Zimowska, G.; Aris, J.P.; Paddy, M.R.
J. Cell Sci. 110, 927-944, 1997
A/Title: A *Drosophila* Tpr protein homolog is localized both in the extrachromal
A/Reference number: Zl1786; MUID:97296455; PMID:9152019
A/Accession: Tl13829
A/Status: preliminary;
A/Molecule type: mRNA
A/Residues: 1-2346 <ZIM>
A/Cross-references: EMBL:U91980; NID:gl9232273; PID:gl9232274; PIDN:AAC47506.
C/Genetics:
A/Cross-references: FlyBase:FBgn0013756
A/Map position: 2R

		7.08; Score 106.5; DB 2; Length 2346;
		Best Local Similarity 18.9%; Pred.No.le+o2;
		Matches 58; Conservative 66; Mismatches 118; Indels 65; Gaps 11
Qy	25 DLNKKYLDVVPKMTFTDEIKELS--RFQEYSQEARSLVGDTKLVLMDSQDYFEATQ	81
	:	:
	:	:
	:	:
Db	203 ELQNIRREHINTMLOCSLCKETESKLWEOVEQAKTIGELTSKIEMNDTAFKNQ	262

QY 82 TVYEWCGVVTQLLSAYILLFDEYNEKAS--AQKDILIRILDDGVKLNKAEQK--SLLTS 137

[illegible][illegible][illegible][illegible][illegible]

.....

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

22

FOR IMMEDIATE RELEASE

RESULT 44

I41078
hemolysin - Escherichia coli
C:Species: Escherichia coli
C:Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 21-Jul-2000
C:Accession: I41078
R:Schmidt, H.; Beutin, L.; Karch, H.
Infect. Immun. 63, 1055-1061, 1995
A:Title: Molecular analysis of the plasmid-encoded hemolysin of Escherichia coli
A:Reference number: I41077; MUID:95172699; PMID:7868227
A:Accession: I41078
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-998 <RES>
A:Cross-references: EMBL:X79839; NID:9860924; PID:CAA56234.1; PID:94388764
C:Superfamily: hemolysin A; hemolysin A homology
C:Keywords: lipoprotein
F:233-776/Domain: hemolysin A homology <HLYA>
F:550.675/Binding site: palmitate (Lys) (covalent) #status predicted

F;233-776/DOMAIN: hemoYsln A hemoYsln <HLIA>
E;550,675/Binding site: palmitate (Lys) (covalent) #status predicted

[illegible]

A;Description: The sequence of S. cerevisiae cosmid 9513.

A;Reference number: S69057

A;Accession: S69058

A;Molecule type: DNA

A;Residues: 1-566 <COU>

A;Cross-references: EMBL:U51033; NID:g1230676; PID:g1230678; MIPS:YPRO70W

C;Genetics:

A;Gene: SGD:MED1

A;Cross-references: SGD:S0006274

A;Map position: 16R

Query Match 6.9%; Score 104.5; DB 2; Length 566;

Best Local Similarity 25.4%; Pred. No. 22;

Matches 57; Conservative 29; Mismatches 91; Indels 47; Gaps 11;

QY 15 SAITAGALDLYNKY-----LDQVPAKTFDETI-KELSRFKQ-----52

DB 6 SYVETLDSMIEFLKDYKPGSTLENITRLCTGLSEFTSELSNELSRSTASKIIVDV 65

QY 53 EYSQEAASVLVDGIKVLMLDSQK--YFEATQTVYEWGCVTQTLSSAYILLFDEYNEKKAS 110

DB 66 DYNKKQD-RIQDVKLVLASNFDFYFNQRDGEHEKSNILLNSLTYPDLKAFHNNLKFL 124

QY 111 AQKILLRILDDGVKKLN-EAKSLTSSQSFNNASGKLLALDLSQLTNDFSEKSSY-----165

DB 125 YLLDAYSHIESDSTSHNNGSSDKSLDSNASFNN-OGKL-----DLFKYFTELSHYIRQC 178

QY 166 FQSQVDRIKAYAGAAAGIVAGPFGLIISYSIAAGVIEGKLIP 209

DB 179 FQDNCCDFKVTNLNDRFGI-----YILTQG-INGKEVP 211

RESULT 48

T48959

N;Alternate names: Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C;Accession: T48959

R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unsel, M.; Mewes, H.W.; Rudd, S.;

submitted to the Protein Sequence Database, April 2000

A;Reference number: 425009

A;Accession: T48959

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1229 <JOR>

A;Cross-references: EMBL:AL163975; GSPDB:GN00061; ATSP:T15B3.190

A;Experimental source: cultivar Columbia; BAC clone T15B3

C;Genetics:

A;Gene: ATSP:T15B3.190

A;Map position: 3

A;Introns: 95/3; 147/3; 175/3; 215/3; 253/3; 284/3; 314/3; 338/2; 363/2; 392/3; 412/2; 4

Query Match

Best Local Similarity 6.9%; Score 104.5; DB 2; Length 1229;

Matches 62; Conservative 55; Mismatches 110; Indels 59; Gaps 13;

QY 2 TSIFAEQVEVVKSAIEADGALDLYNKYLDOVPWKTFTDTIKELSRFKQ-EYSQEAASV 60

DB 967 SAVVFQOEERARAHVNSFGYLQKQNEELDVIKSHKR--EIDAMGKIQQSEAEKLSNI 1024

QY 61 LVGDIKVLMLDSQKYPFATQTVYEWG-----GVVITQLLSAY-----ILFDEY-----NE 106

DB 1025 VMLKIKV--DEENKRHEEGVL--CTIDNILRTGRATDLKQSEKTKLOSEMKLRSRE 1078

QY 107 KASNAQKILLRILDDGVKKLNEAKSKLITSSQSFNNASGKLLALDLSQLTNDFSEKS---163

DB 1079 KLASVRKEV-----DDMTKSLKUEKEIKTWEIETKSSKTRTSEMELENTIQEROTIQ 1133

QY 164 -----SYFQSDVRI-----RKEAYAGAAAGIVAGPFGLIISYSIAAGVIEGKL 207

DB 1134 EMEBQGMSEIQNWITEIHLVFESDLRKE-----EAMIRRE--LTAEBELRAKDVTHTM 1185

QY 208 IPELANNRLKTYQNFPTLSATVKQANKDIDAAKLKLATEIAAIGEI 253

DB 1186 IERVENALKTLENQNNNSVGKIEEVENV-----LSLVHEASRLLEV 1227

RESULT 49

VMUT21

VSG expression site-associated protein 221a precursor - Trypanosoma brucei

N;Alternate names: ESAG protein

C;Species: Trypanosoma brucei

C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 16-Jul-1999

C;Accession: A03395

R;Cully, D.F.; Ip, H.S.; Cross, G.A.M.

Cell 42, 173-182, 1985

A;Title: Coordinate transcription of variant surface glycoprotein genes and an expressi

A;Reference number: A90888; MUID:85254917; PMID:2861910

A;Molecule type: mRNA

A;Residues: 1-329 <CU>

A;Cross-references: GB:M11452; NID:g162072; PID:AAA30191.1; PID:g162073

C;Comment: The function of the ESAG proteins is not known but may be related to activat

C;Keywords: glycoprotein

F;2-23/Domain: signal sequence #status predicted <SIG>

F;24-329/Product: VSG expression site-associated protein 221a #status predicted <MAT>

F;73,294,308/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.9%; Score 104; DB 1; Length 329;

Best Local Similarity 21.9%; Pred. No. 11;

Matches 46; Conservative 34; Mismatches 64; Indels 66; Gaps 9;

QY 92 QLLSAYILLFD-----EYNEKASAKQDILIRILDGVKKL-NEAOKS 133

DB 4 EIVELWLLFVSVCVDLAWLQCADCTRVADHKEHAPVTEAVCYLCRLSLDALNKLKLYSEGEKK 63

QY 134 LLTSSQSFNNASGKLLALDLSQLTNDFSEKSSY-----FOSQVDRIKAYAGAAAGIV 186

DB 64 LLVTEVYANAS---LILD-DWEGRAGESSTYLSVIRGVMEQTDRLK-----108

QY 187 AGPFGLIISYSIAAGVI---EGKLIPELNNRLKTVQ-----NFFTSLSATVKQA 232

DB 109 -----LISYGNKMGMLVAKAGGLPALEDSLKEVKEIPGALIKTNKYTTSVAEIVRTV 162

QY 233 NKDIDAANKL-----ATEIAAIGEIKTE 256

DB 163 WEDVGEILLWKETEAKCGSQKVEGVGEIOTE 192

RESULT 50

T41332

cap homolog - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999

C;Accession: T41332

R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.

submitted to the EMBL Data Library, March 1998

A;Reference number: Z21988

A;Accession: T41332

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-633 <WOO>

A;Cross-references: EMBL:AL022243; PID:CAA18286.1; GSPDB:GN00067; SPDB:SPCC364.04c

A;Experimental source: strain 972h; cosmid c364

C;Genetics:

A;Gene: SPDB:SPCC364.04c

A;Map position: 2 55/1

A;Introns: 18/1; 55/1

Query Match

Best Local Similarity 6.9%; Score 104; DB 2; Length 633;

Matches 81; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

QY 38 KTFDETIKLSRFKQEY-SQEAASVLVDGIKVL-----MDSQDKYFEATQT-----VYEW 86

```

Db      40 KTSLEDERKELSKTEPRKQDPDEVKLGEMKGLLKYQSGIDSLTKRAKSABATFRVYET 99
QY      87 CGVVTQLLSAYILLFDEYNEKASIQ-----KDILIRILDDGVKKLN- 128
Db      100 LGEVP---DPYPLLIJEAANNLTKQIEDLKKKEEMEGSLQGGKELEREVENLRKELDK 156
QY      129 -----EAKSLITSSQSF-----NNAS-----GKLLALD-SQLTND 158
Db      157 YKDLVETEAEKAAITKECEKSWLEQQKLYKDMEQENASTIQKTSKIRELQASQLDHD 216
QY      159 FSEKSYFOSQVDRIRKAYAGAAAGIVAGPPGLIISYSIAAGVIE-----GKLIPE 210
Db      217 L-QASQNESAGLDVNAKSAEVNAILSELDDANKIIVELQAEIATVLKONTKEQKSGSQDD 275
QY      211 LNNRLKTVONFTTSLSATVKQANKDI-DAAKLKLATEIAAICEIKETETETTR----- 261
Db      276 LSNQKQQLDFMDSLNNKKLSTELSIKEASRKEMETHCATIQTLENEVKARKYKESLT 335
QY      262 ----FYVDYDDLK--LSLLK-----GAAKKMINTCNEYQORHGKKTLE 299
Db      336 LANKPSDYDEIKRELSVLKQIEFSGEHATHENTSIQSLLKREKOLSE 363

```

Search completed: November 28, 2003, 13:50:21
 Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2003, 13:48:16 ; Search time 17 Seconds
(without alignments)
843.714 Million cell updates/sec

Title: US-09-993-292A-2
Perfect score: 1515
Sequence: 1 MTSIFAQTEVWKSIAETA.....NEYQRHGKKTLPFVDPVAS 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1402	92.5	303	1	HLV1_ECOLI
2	116	7.7	1023	1	P77335 escherichia
3	115.5	7.6	622	1	S688_CAEEL
4	113	7.5	956	1	YEP3_YEAST
5	113	7.5	1005	1	RA50_METUA
6	113	7.5	1628	1	NAGH_CLOPE
7	112.5	7.4	1024	1	HLV1_ECOLI
8	112	7.4	1102	1	MYSC_CHICK
9	110.5	7.3	284	1	TPM_MYTED
10	110.5	7.3	284	1	TPM_MYTGA
11	110.5	7.3	584	1	SVY_SVNY3
12	110	7.3	1338	1	MYSD_CAEEL
13	110	7.3	2330	1	G034_HUMAN
14	109	7.2	955	1	LKAA_PASHA
15	108.5	7.2	1066	1	KL61_DROME
16	107.5	7.1	1875	1	MLP1_YEAST
17	106.5	7.0	1500	1	SSP5_STRGN
18	106	7.0	953	1	LK33_PASHA
19	104	6.9	329	1	ES02_TRYBB
20	104	6.9	2334	1	WAPA_BACSU
21	103.5	6.8	520	1	RECN_AQUAE
22	103.5	6.8	2017	1	MYSN_DROME
23	103	6.8	756	1	Y328_MYCGE
24	103	6.8	978	1	RA50_AQUAE
25	103	6.8	1726	1	MSP1_PLAPP
26	103	6.8	5171	1	BPEA_HUMAN
27	102.5	6.8	2022	1	ANT1_ONCVU
28	102	6.7	520	1	H855_RICPR
29	102	6.7	1376	1	MYHA_BOVIN
30	102	6.7	1376	1	MYHA_HUMAN
31	101.5	6.7	982	1	P115_MYCGE
32	101.5	6.7	3680	1	DMD_CANFA
33	101	6.7	794	1	HMMR_MOUSE

RESULT 1

ALIGNMENTS

101	1403	6.7	1403	1	VG22_HSVII
101	1398	6.7	1398	1	MYH6_MOUSE
101	1398	6.7	1398	1	MYH6_HUMAN
100.5	457	6.6	457	1	MESE_LEUME
100.5	643	6.6	643	1	K2C1_HUMAN
100.5	643	6.6	643	1	TEA1_SCHPO
100.5	1147	6.6	1147	1	SMC2_MOUSE
100.5	1191	6.6	1191	1	SBC_LACLA
100	1466	6.6	1466	1	MYH9_HUMAN
100	1860	6.6	1860	1	DESP_HUMAN
100	2871	6.6	2871	1	MYH9_CHICK
99.5	1859	6.6	1859	1	MYH9_RAT
99.5	1976	6.6	1976	1	RBPI_PLAVB
99.5	2869	6.6	2869	1	MSP1_PLAVB
99	1726	6.5	1726	1	MAFE_HUMAN
98.5	5430	6.5	5430	1	SNEI_HUMAN
98.5	8797	6.5	8797	1	RA50_THEMA
98	852	6.5	852	1	KF14_HUMAN
98	1648	6.5	1648	1	MYH6_RAT
98	1938	6.5	1938	1	MYH6_MESAU
98	1939	6.5	1939	1	DMD_MOUSE
98	3678	6.5	3678	1	DYHA_CHLRE
98	4899	6.5	4899	1	SYFB_BACHD
97.5	808	6.4	808	1	USO1_YEAST
97.5	1509	6.4	1509	1	MYH8_HUMAN
97.5	1790	6.4	1790	1	CENF_HUMAN
97.5	3210	6.4	3210	1	CDG3_YEAST
97	520	6.4	520	1	VDP_BOVIN
97	961	6.4	961	1	SMC3_SCHPO
97	1194	6.4	1194	1	MSP1_PLAFF
97	1701	6.4	1701	1	MSP1_PLAFM
97	265	6.4	265	1	APAI_TUPGB
96.5	1197	6.4	1197	1	SMC2_HUMAN
96.5	1941	6.4	1941	1	MYH2_HUMAN
96.5	2349	6.4	2349	1	TPR_HUMAN
96.5	3685	6.4	3685	1	DMD_HUMAN
96	321	6.3	321	1	EABE_ECO27
96	351	6.3	351	1	TRPF_LACLA
96	714	6.3	714	1	CLPB_MYCGE
96	877	6.3	877	1	DPO1_LACLC
96	975	6.3	975	1	KINH_DROME
96	1130	6.3	1130	1	VL17_CAEEL
96	1361	6.3	1361	1	MYH9_RAT
95.5	1052	6.3	1052	1	RPOC_BACAN
95.5	1939	6.3	1939	1	MYH1_HUMAN
95	613	6.3	613	1	PCD8_HUMAN
95	1163	6.3	1163	1	SBC_CLOAB
94.5	265	6.2	265	1	YB46_MYCPN
94.5	1041	6.2	1041	1	UN83_CAEEL
94	576	6.2	576	1	RECN_BACSU
94	609	6.2	609	1	GLMS_CLOPE
94	863	6.2	863	1	MYSP_ECHGR
94	1088	6.2	1088	1	RRPO_ROT1
94	1111	6.2	1111	1	KIP1_YEAST
94	1193	6.2	1193	1	LMG2_HUMAN
94	1935	6.2	1935	1	MYH7_HUMAN
93.5	558	6.2	558	1	YAB3_SCHPO
93.5	727	6.2	727	1	MEP1_ARATH
93.5	928	6.2	928	1	KINH_NEUCR
93.5	1013	6.2	1013	1	AG0D_DROME
93.5	1111	6.2	1111	1	KIP1_YEAST
93.5	1526	6.2	1526	1	MYG2_SCHPO
93	364	6.1	364	1	MURG_THETN
92.5	284	6.1	284	1	TPM1_MOUSE
92.5	284	6.1	284	1	TPM1_RABIT
92.5	284	6.1	284	1	TPM1_RAT
92.5	284	6.1	284	1	TPM2_BIOGL

ALIGNMENTS

HLVE_ECOLI
 ID HLVE_ECOLI STANDARD; PRT; 303 AA.
 AC P7735; Q47276; Q9R3G4;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemolysin E, chromosomal (Hemolysis-inducing protein) (Silent
 DE hemolysin sheA) (latent pore-forming 34 kDa haemolysin).
 GN HLVE OR HPR OR SHEA OR CLYA OR B1182.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / XLI-BLUE;
 RA McNamara P.J., Iandolo J.J., Unlich G.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA del Castillo F.J., Leal S.C., Moreno F., del Castillo I.;
 RL "The Escherichia coli K-12 sheA gene encodes a 34-kDa secreted
 RT haemolysin."
 RL Mol. Microbiol. 25:107-115(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA Ludwig A., Bauer S., Benz R., Goebel W.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3030-2;
 RA Xing J., Fernandez S.V., Kapur V., Barletta R.G., Moxley R.A.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RL "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RL "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155(1996).
 RN [7]
 RP SEQUENCE OF 1-296 FROM N.A.
 RC STRAIN=K12 / XLI-BLUE;
 RA King C.H., Shinnick T.M.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE OF 1-156 FROM N.A.
 RC STRAIN=K12 / AB1157;
 RA Woodgate R.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=20123443; PubMed=10650049;
 RA Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A.,
 RA Green J., Artymiuk P.J.;

"E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of
 the toxin and observation of membrane pores by electron microscopy."
 Cell 100:265-276(2000).
 CC -|- FUNCTION: HEMOLYTIC ON SHEEP BLOOD AGAR. FORMS PORE.
 CC -|- SUBUNIT: HOMOLIGOMERIZE.
 CC -|- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -|- DOMAIN: THE TRANSMEMBRANE DOMAIN IS BELIEVED TO BE INVOLVED IN
 CC PORE FORMATION BY THE CYTOTOXIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U57430; BAB07048.1; ALT INIT.
 DR EMBL; X98615; CAA67204.1; ALT_INIT.
 DR EMBL; AJ001829; CAA05035.1; -.
 DR EMBL; U73842; AAD04731.1; -.
 DR EMBL; AE000216; AAC74266.1; -.
 DR EMBL; D90751; BAA36016.1; ALT_INIT.
 DR EMBL; D90752; BAA36029.1; ALT_INIT.
 DR EMBL; D90753; BAA36037.1; ALT_INIT.
 DR EMBL; U22466; AAA92081.1; -.
 DR EMBL; U13610; -; NOT_ANNOTATED_CDS.
 DR PDB; 1QOY; 23-JAN-00.
 DR EcoGene; EG13243; hlyE.
 DR Hemolysis; Toxin; Transmembrane; Cytolysis; Complete proteome;
 KW 3D-structure.
 FT TRANSMEM 81 101 POTENTIAL.
 FT HELIX 1 28
 FT TURN 29 29
 FT HELIX 30 34
 FT HELIX 37 46
 FT TURN 47 50
 FT HELIX 51 53
 FT HELIX 56 99
 FT TURN 100 103
 FT HELIX 106 159
 FT TURN 161 162
 FT HELIX 164 179
 FT TURN 180 184
 FT STRAND 185 187
 FT HELIX 189 191
 FT STRAND 193 195
 FT HELIX 196 199
 FT TURN 200 201
 FT TURN 204 205
 FT HELIX 207 258
 FT TURN 259 259
 FT HELIX 268 291
 FT TURN 292 293
 SQ SEQUENCE 303 AA; 33758 MW; 9BE348DA095668A5 CRC64;
 Query Match 92.5%; Score 1402; DB 1; Length 303;
 Best Local Similarity 91.1%; Pred. No. 2.7e-87;
 Matches 276; Conservative 16; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MTSIFAQETVEVWKSATETADGALDLYNKYLDQVTPWKTDETIKELSRFKQEYSQASV 60
 DB 1 MTEIVADKTVEVWNAIETADGALDLYNKYLDQVTPWQTDETIKELSRFKQEYSQASV 60
 QY 61 LVGDIKLLMDSQDKYFEATQTVYEWGCVTQLLSAVILLFDEYNEKKAQAQDILIRIL 120
 DB 61 LVGDIKLLMDSQDKYFEATQTVYEWGCVATQLLAAYILLFDEYNEKKAQAQDILIKVL 120
 QY 121 DGVKKLNEAOKSLTSSQSPNNASGKLALDLSOLTNDPSEKSYFOSQVDRIRKEAYAG 180
 DB 121 DGGITKLENAQKSLVSSQSPNNASGKLALDLSOLTNDPSEKSYFOSQVDRIRKEAYAG 180
 QY 181 AAAGIVAGFFGLIIISYSIAAGVIEGKLIPELNRLKTVQNFVFFSLSATVQKANKDIDAAK 240

Db 181 AAAGVAGPGLIISYIAAGVWEGKLPKLNKLSKVQVFFTLSTNTVQANKDIDAAK 240
QY 241 LKLAATEAAGIEIKTETTRFYVDYDMLSLKGAAGKMINTCNEYQORHGKTLFEV 300
Db 241 LKLAATEAAGIEIKTETTRFYVDYDMLSLKGAAGKMINTCNEYQORHGKTLFEV 300
QY 301 PDV 303
Db 301 PEV 303

RESULT 2

HLV1_ECOLI STANDARD; PRT; 1023 AA.
ID HLV1_ECOLI STANDARD; PRT; 1023 AA.
AC P09983;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemolysin, chromosomal.
GN HLVA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J96 / Serotype O4;
RX MEDLINE=85234404; PubMed=3891743;
RA Feilmeier T., Pellett S., Welch R.A.;
RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";
RL J. Bacteriol. 163:94-105(1985).
RP SEQUENCE OF 1-44 FROM N.A.
RC STRAIN=2001;
RX MEDLINE=85258115; PubMed=3894051;
RA Nicoud J.-M., Mackman N., Gray L., Holland I.B.;
RT "Characterisation of HlyC and mechanism of activation and secretion
of haemolysin from E. coli 2001.";
RL FEBS Lett. 187:339-344(1985).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
cell membranes and cause cell rupture by mechanisms not clearly
defined.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: The Gly-rich region is probably involved in binding
calcium, which is required for target cell-binding or cytolytic
activity.
CC -!- DOMAIN: The three transmembrane domains are believed to be
involved in pore formation by the cytotoxin.
CC -!- PTM: Palmitoylated by hlyC. The toxin only becomes active when
modified.
CC -!- MISCELLANEOUS: The hemolysin of E.coli is produced predominantly
by strains causing extraintestinal infections, such as those of
the urinary tract.
CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; M10133; AA023975.1; -;
CC EMBL; X02768; CA026546.1; -;
CC PIR; A24433; LEECA.
CC InterPro; IPR001343; Hemlyen_Ca_bind.
CC InterPro; IPR003995; RTXa.
CC Pfam; PF00353; hemolysinCabd; 6.
CC Pfam; PF02382; RTX; 1.
CC PRINTS; PR00313; CABDNCRPT.
CC PRINTS; PR01488; RTXTOXINA.

DR PROSITE; P500330; HEMOLYSIN_CALCIUM; 4.
KW Hemolysis; Toxin; Cytolysis; Repeat; Calcium; Transmembrane;
Lipoprotein; Palmitate.
FT TRANSMEM 237 259 POTENTIAL.
FT TRANSMEM 267 326 POTENTIAL.
FT TRANSMEM 364 410 POTENTIAL.
FT DOMAIN 723 869 16 X REPEATS, GLY-RICH.
FT REPEAT 723 728 1.
FT REPEAT 732 737 2.
FT REPEAT 741 746 3.
FT REPEAT 750 755 4.
FT REPEAT 759 764 5.
FT REPEAT 768 773 6.
FT REPEAT 777 782 7.
FT REPEAT 786 791 8.
FT REPEAT 795 800 9.
FT REPEAT 806 812 10.
FT REPEAT 816 821 11.
FT REPEAT 825 830 12.
FT REPEAT 834 839 13.
FT REPEAT 843 848 14.
FT REPEAT 855 860 15.
FT REPEAT 864 869 16.
FT LIPID 563 563 PALMITATE (BY SIMILARITY).
FT LIPID 689 689 PALMITATE (BY SIMILARITY).
FT VARIANT 6 6 A -> T (IN STRAIN 2001).
SQ SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64;
Query Match 7.7%; Score 116; DB 1; Length 1023;
Best Local Similarity 20.5%; Pred. No. 2.6; Indels 98; Gaps 11;
Matches 69; Conservative 59; Mismatches 111;
QY 2 TSFAEQTVVEVVK-----SAIETADGALDLYNKYLDQVTPKTFDTIKELSRPQK 53
Db 155 TALSSMKIDELIKKQSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFSQ 207
QY 54 YQEQASVL-----VGDIKVLLMSQDKYFEATQTVYVCGVVTQLLSAYILL-FDEY 104
Db 208 LNKGLSVLNTKHLNGVGN-KLQNLPLNDNTGAGLDTV---SGILSALSASFILSNAD 263
QY 105 NEKKSASQKIDILIRLDD---GVKKLENAQKSLTSSQFNNSAGKLLALDSQLTNDPSE 161
Db 264 TGTAAAGVELTTKVLGVNGKISQVIAAQAAGLSLTS----- 302
QY 162 KSGYFQSQVDRIKRAYAGAAAGVAGPGLIIS-----YSTAAGVIEGKLPDLNRLKT 217
Db 303 -----AAAAGLIASVVTLAISPLSLFIADKFRANKIEYSQRFKK 344
QY 218 VQNFFTSLSATVKQANKDIDAAKLAATEAAGIEIKTETTRFYVDYDMLSLKGA 277
Db 345 LGYDGSLLAAATHKETGATGIDASLSTVLSVSSGISAATTSLV-----GAPVSAALYGA 400
QY 278 -----AKGMINTONEYOORHGK 294
Db 401 VTGIISGILEASKQAMFHVASKMADVAIEWEKKGK 437
RESULT 3
SR68_GABEL STANDARD; PRT; 622 AA.
ID SR68_GABEL AC Q20822;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable signal recognition particle 68 kDa protein (SRP68).
GN F55C5.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
CC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;

```
RA Harris B.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC RETICULUM
CC MEMBRANE. SRP68 BINDS THE 7S RNA. SRP72 BINDS TO THIS COMPLEX
CC SUBSEQUENTLY. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT INTERACT
CC DIRECTLY WITH THE DOCKING PROTEIN IN THE ER MEMBRANE AND POSSIBLY
CC PARTICIPATE IN THE ELONGATION ARREST FUNCTION (BY SIMILARITY).
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SRP68 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z78198; CAB01573.1; -.
CC PIR; T22716; T22716.
CC DR Wormpep; F55C5.8; CE20875.
CC KW Hypothetical protein; Signal recognition particle; Ribonucleoprotein;
CC RNA-binding.
CC SQ SEQUENCE 622 AA; 70574 MW; A7B8808E46169636 CRC64;
Query Match 7.6%; Score 115.5; DB 1; Length 622;
Best Local Similarity 21.2%; Pred. No. 1.5;
Matches 69; Conservative 56; Mismatches 109; Indels 87; Gaps 15;
QY 41 DETIKELSRPKQYSQASVLDGDKVLLMSQDKYFEATQTVYEWCGVVTQLLSAY--- 97
DB 251 DKLISEM-RASATSAEVAWVIEWGAKSTVDDE-----KAKQVQEWKQTEVELAQCPKP 304
QY 98 --ILLFDEYNEKASAKOILIRLDDGVKKLNE-AQKSLTSSQGF-----NNASGKL 148
DB 305 EKVALP---EKATADTRDAIDRSIDIRKSSSENADTTVLQSIKAYLEFLKMGNTASRY 360
QY 149 LALDSOLTDNDFSKSYFOS-----IDNTKSEKSKFPDLLYDSVTEIYKEVAEIPGADHKNLIDAEFVKEVYRA 417
DB 361 LAI---IDNTKSEKSKFPDLLYDSVTEIYKEVAEIPGADHKNLIDAEFVKEVYRA 417
QY 181 AAGAGVAGPGLIISYSIAAGVI-----EGKLIPELNRLKTVQNFSTSATVK 230
DB 418 FRCFYMASSYSALHKYSERAALEDRTVSRVQDAEGKL-----KKLKSSSFITNETQSSLN 472
QY 231 QANKDIDAAK--LKLATEIAAICEIKTETETTFYVDYDMLSLKGAKKWINTCNEY 288
DB 473 ELSEVESAKTVRAARLASAGDVKTDSELAKI-ID-----KRPLETFVNEW 519
QY 289 QQ---RHGKKTLPFEPDVAS 305
DB 520 RQWDVRNSLKKTKIP-VAS 538
RESULT 4
YEF3_YEAST STANDARD; PRT; 956 AA.
AC P32618;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 106.1 kDa protein in GLY1-GDAl intergenic region.
GN YEL043W OR SYGP-ORF14.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=S288c / AB972;
RX PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA Oh C., Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL Nature 387:78-81(1997).
RN [2]
RP FIBRONECTIN TYPE III DOMAIN.
RX MEDLINE=97148176; PubMed=8994808;
RA Bateman A., Chochia C.;
RT "Fibronectin type III domains in yeast detected by a hidden Markov
RT model.";
RL Curr. Biol. 6:1544-1546(1996).
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18779; AAB64999.1; -.
CC PIR; S30834; S30834.
CC DR SGB; S0000769; YEL043W.
CC DR InterPro; IPR003961; FN_III.
CC DR SMART; SM00060; FN3; 1.
CC KW Hypothetical protein.
CC FT DOMAIN 35 125 FIBRONECTIN TYPE-III.
CC SQ SEQUENCE 956 AA; 106132 MW; 3F78B09A0FCA03AF CRC64;
Query Match 7.5%; Score 113; DB 1; Length 956;
Best Local Similarity 24.4%; Pred. No. 3.8;
Matches 68; Conservative 45; Mismatches 78; Indels 88; Gaps 15;
QY 38 KTFDETIKELSRPKQYSQASVLDGDKVLLMSQDKYFEATQTVYEWCGVVTQ 92
DB 272 KSLKSNIKSL-----ENSKLLDTLTKLKKDKSEKISKWENDMQKMSQEDTE 322
QY 93 LLS-----AYILLFDEYNEKASAKOILIRLDDGVKKLNEAKSLTS---S 138
DB 323 LLSKDTTIKERYFKLLNESNASVANINKIEISLQNEISKMEESKRLNASKSLITSVN 382
QY 139 QSFNN---ASGKLALDSOLTDNDFS-EKSSYF-----QSQVDRIRKEAYA 179
DB 383 ANVENDPIASGELSALVKKL-NDFTLEKNGFLSNAGEEFLSKLNADSSLIKMKQB--- 438
QY 180 GAAAGVAGPGLIISYSIAA-----GVIEGKLIP-ELNNR-LKT-----V 218
DB 439 -----LSIDQELANWKLQSRNLLKKISALENQFNWMSLNRLNLTCLMAYQPY 486
QY 219 QNFFTSLSAT-----VKQANKDIDAAKKLATEIAAIGEI 253
DB 487 KNGDLSAATNSNNSAEKRSKSSGSIQLPLSNNSRTGSI 525
RESULT 5
RA50_METJUA STANDARD; PRT; 1005 AA.
ID RA50_METJUA
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 Atpase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
```

OC Methanocaldococcaceae; Methanocaldococcus.
RN NCBI_TaxID=2190;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Huxst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser K.M., Smith H.O., Woese C.R., Venter J.C.,
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL Science 273:1058-1073 (1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67572; AAB99331.1; -.
CC TIGR; M31322; -.
CC HAMAP; MF_00449; -; 1.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR003405; SMC C.
CC InterPro; IPR003395; SMC N.
CC Pfam; PF04423; Rad50 zn Hook; 1.
CC Pfam; PF02483; SMC_C_1.
CC Pfam; PF02463; SMC_N_1.
CC ProDom; PD000006; ABC transporter; 1.
CC SMART; SM00382; AAA; 1.
CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
KW NP BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).
SQ SEQUENCE 1005 AA; 119387 MW; 9BB3B48173E788F3 CRC64;

Query Match 7.5%; Score 113; DB 1; Length 1005;
Best Local Similarity 19.7%; Pred. No. 4;
Matches 60; Conservative 55; Mismatches 91; Indels 98; Gaps 12;

QY 7 EOTVEVWKAETADGALDLYNKLDQVLPKTFDE-----TIKELSRF-----KOEYSQ 56
DB 605 DILIEDIKQLNKKF--NFYQYLSAVSYLNSVDDEGIRIKKIEINIVSGWKKKRE 661
QY 57 EASVL-----VGDIKVLMSQDKYFEATQVYEWGCVTQLLSAYILLDFEYNEKKAS 110
DB 662 ELNKLREDBREINRLKDLNKLNKKKE-----LIEIENRRSLKFDKYEYLG 710
QY 111 AOKDILIRLDGVKVKNEAQSJLTSSQSFNNASGKLLAD----- 152
DB 711 TKELRLKNKDKGLBEI-----YNICMSKILADNIRKYNKSDIEIYLNK 757
QY 153 -----SOLTNDPFEKSYFSQVDRIRKEAYAGAAAGVAGPFGLIISYIAAGVIEGKLI 208
DB 758 ILEVNKEINDEIRISYINQKLE-----INVEE-----EHKKI 792
QY 209 PEL--NNR--LXTQVNFSTLSATVQKQNDIAKDKLKLATLAEIAIGIKETETTFYV 264

Db 793 KELYENKROELDNVRQKTEITGIEYLYKKDVESIKARL-----KEMSNLEKEKELTKEV 848
QY 265 DYDD 268
Db 849 EYLD 852

RESULT 6
NAGH CLOPE STANDARD; PRT; 1628 AA.
ID NAGH CLOPE STANDARD; PRT; 1628 AA.
AC P26831;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hyaluronoglucosaminidase precursor (EC 3.2.1.35) (Hyaluronidase)
DE (Mu toxin).
GN NAGH OR CPE0191.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=94232189; PubMed=8177218;
RA Canard B., Garnier T., Saint-Jeanis B., Cole S.T.;
RT "Molecular genetic analysis of the nagH gene encoding a hyaluronidase
RT of Clostridium perfringens";
RL Mol. Gen. Genet. 243:215-224 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
CC -!- FUNCTION: Putative virulence factor which is likely to act on
CC connective tissue during gas gangrene.
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1->4-linkages between N-
CC acetyl-beta-D-glucosamine and D-glucuronate residues in
CC hyaluronate.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 F5/8 type C domain.
CC -!- CAUTION: The partially purified protein from strain CPN50 is
CC approximately 70 kDa smaller than the sequence indicated here.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81878; AAA3259.1; -.
CC EMBL; AP003185; BAB79897.1; -.
CC PIR; S43904; S43904.
CC InterPro; IPR002105; Dockerin_1.
CC InterPro; IPR000421; FAS8 C.
CC Pfam; PF00404; Dockerin_1; 2.
CC Pfam; PF00754; F5_F8 type_C; 1.
CC SMART; SM00231; FAS8C_1.
CC PROSITE; PS00022; FAS8C_1.
KW Hydrolase; Glycosidase; Toxin; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1628 HYALURONOGUCOSAMINIDASE.
FT DOMAIN 781 953 F5/8 TYPE C.
FT VARIANT 147 147 G -> A (IN STRAIN CPN50).
FT VARIANT 172 175 KGS -> EIKN (IN STRAIN CPN50).
FT VARIANT 250 250 V -> M (IN STRAIN CPN50).

FT VARIANT 548 548 A -> E (IN STRAIN CPN50).
 FT VARIANT 558 558 D -> E (IN STRAIN CPN50).
 FT VARIANT 614 614 G -> S (IN STRAIN CPN50).
 FT VARIANT 944 944 I -> V (IN STRAIN CPN50).
 FT VARIANT 950 950 N -> S (IN STRAIN CPN50).
 FT VARIANT 979 979 T -> I (IN STRAIN CPN50).
 FT VARIANT 982 982 I -> L (IN STRAIN CPN50).
 FT VARIANT 1042 1042 I -> F (IN STRAIN CPN50).
 FT VARIANT 1043 1043 MISSING (IN STRAIN CPN50).
 SQ SEQUENCE 1628 AA; 182474 MW; D4252A2512BBED69 CRC64;

Query Match 7.5%; Score 113; DB 1; Length 1628;
 Best Local Similarity 23.2%; Pred. No. 7.3;
 Matches 72; Conservative 42; Mismatches 112; Indels 84; Gaps 14;

QY 27 YNKYLDQVPMKTFDETIKELSRFKQVSYQSEAS-VLVGDIKVLMLDSQDKYFEATQTVYE 85
 DB 1328 YNNYIEKIETWDVAFNIYEITLNOEFEPVNDLSLKAKYDELINLSGDEV----- 1378

QY 86 WCGVVTQLLSAYILLDFEYNEKA-----SAQKILIRILDGKVKNEAKSL---LT 136
 DB 1379 -----TLSSFTLKEALNEAKSILDDNSQKKI-----DKALEKLNKAEERLDLRAT 1426

QY 137 SSQSFNNASGKLALDSQLTND-----FSE-----KSSYFQSQVDRIKKE 176
 DB 1427 DFEDFN-----KVLTLGNSLVEEYTAESWALFSEVLEANEANKKADYTQDQINQIVID 1482

QY 177 AYAGAAGIVAGP-----FGLIISY--STAAGVIEGKLIPE-----LNNRLKTVQN 220
 DB 1483 LDASIKALVKETPEVDKTNLGLINGQSLDESVEGFNVGHYKAKDGLVEINKAAEE 1542

QY 221 FETLSATVKQAKNDIDAKKLATIAAIGETKTETTRF-----YVDYDQLMLSLKLG 276
 DB 1543 VFKEKDAT-----EEINLAKESLEGALAFNSLLIEESIGDFNGNKIDIGDLAM----- 1593

QY 277 AAKMINTON 286
 DB 1594 VSKNIGSTTN 1503

RESULT 7

HLVA_ECOLI STANDARD; PRT; 1024 AA.
 AC P08715;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemolysin, plasmid.
 GN HLVA.
 OS Escherichia coli.
 OG Plasmid Inci2 PHLY152.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hess J., Wells W., Vogel M., Goebel W.;
 RT "Nucleotide sequence of a plasmid-encoded hemolysin determinant and
 its comparison with a corresponding chromosomal hemolysin sequence.";
 RL FEMS Microbiol. Lett. 34:1-11(1986).
 RN [2]
 RP PALMITOYLATION OF LYS-564 AND LYS-690.
 RX MEDLINE=5039325; PubMed=7801126;
 RA Stanley E., Packman L.C., Koronakis V., Hughes C.;
 RT "Fatty acylation of two internal lysine residues required for the
 toxic activity of Escherichia coli hemolysin.";
 RL Science 266:1992-1996(1994).
 RN [3]
 RP PALMITOYLATION OF LYS-564 AND LYS-690.
 RX MEDLINE=96404790; PubMed=8808931;
 RA Ludwig A., Garcia F., Bauer S., Jarchau T., Benz R., Hoppe J.,
 RA Goebel W.;
 RT "Analysis of the in vivo activation of hemolysin (HlyA) from

Escherichia coli.";
 J. Bacteriol. 178:5422-5430(1996).
 CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
 cell membranes and cause cell rupture by mechanisms not clearly
 defined.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DOMAIN: The Gly-rich region is probably involved in binding
 calcium, which is required for target cell-binding or cytolytic
 activity.
 CC -!- DOMAIN: The three transmembrane domains are believed to be
 involved in pore formation by the cytotoxin.
 CC -!- PTM: Palmitoylated by hlyC. The toxin only becomes active when
 modified.
 CC -!- MISCELLANEOUS: The hemolysin of E.Coli is produced predominantly
 by strains causing extraintestinal infections, such as those of
 the urinary tract.
 CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M14107; AAA98233.1; --
 DR InterPro; IPR001343; Hemlysn_Ca_bind.
 DR InterPro; IPR003995; RtxA.
 DR Pfam; PF00353; hemolysincabind; 6.
 DR Pfam; PF02382; RTX; 1.
 DR PRINTS; PR00313; CABNDNGRPT.
 DR PRINTS; PR01488; RYTXOXINA.
 DR PROSITE; PS00330; HEMOLYSIN_CALCIUM; 4.
 KW Hemolysin; toxin; Cytolysis; Repeat; Calcium; Transmembrane;
 KX Lipoprotein; Palmitate; Plasmid.
 FT TRANSMEM 238 260 POTENTIAL.
 FT TRANSMEM 268 327 POTENTIAL.
 FT DOMAIN 724 870 16 X REPEATS, GLY-RICH.
 FT REPEAT 724 729 1.
 FT REPEAT 733 738 2.
 FT REPEAT 742 747 3.
 FT REPEAT 751 756 4.
 FT REPEAT 760 765 5.
 FT REPEAT 769 774 6.
 FT REPEAT 778 783 7.
 FT REPEAT 787 792 8.
 FT REPEAT 796 801 9.
 FT REPEAT 807 812 10.
 FT REPEAT 817 822 11.
 FT REPEAT 826 831 12.
 FT REPEAT 835 840 13.
 FT REPEAT 844 849 14.
 FT REPEAT 856 861 15.
 FT REPEAT 865 870 16.
 FT LIPID 564 564 PALMITATE.
 FT LIPID 690 690 PALMITATE.
 SQ SEQUENCE 1024 AA; 110201 MW; 83944917F76C945B CRC64;

Query Match 7.4%; Score 112.5; DB 1; Length 1024;
 Best Local Similarity 20.8%; Pred. No. 4.5;
 Matches 70; Conservative 55; Mismatches 115; Indels 97; Gaps 11;

QY 2 TSIFAEQTVVVK-----SAIETAGDALDLYNKYLDQVPMKTFDETIKELSRKQE 53
 DB 155 TALSSMKIDELIKKQSGNVSSSELAKASIELINQLVDTVASLN-----NNVNSFSQ 208

QY 54 YSQEASVL-----VGDIKVLMSQDKYFEATQVYEWCGVVTOLLGAYILL-FDEY 104
 DB 209 LNTVGSVLNTHNGVGN-KLQNPDLNIGAGLDIV---SGILSAISAFILSNADAD 264

QY 105 NEKASAKQDILIRILD--GVKXKLENAQKSLITSSQSFNNASGKLALDSQLTNDFSE 161

Db 265 TRTKAAAGVELTTKVLGVNGKISQYIIAQRAGLSTSS----- 303
 Qy 162 KSSYFQSQVDRRKREAYAGAGIVAGPFGILIS-----YSTAAGVIEGKLIPELNNRLKT 217
 Db 304 -----AAGLAGASAVTTAISPISLSLADFKRANKIEEYSQRFKK 345
 Qy 218 VQNFPTSLSATVQKQKIDRAKLKATEIAAIGBKITETETTFYVDYDDLMLSLKGA 277
 Db 346 LQYDGDLSLLAAPHKETGAIDASLTITISVLASVSGISAAATTSILV-----GAPVSALVGA 401
 Qy 278 -----AKKMINTCNEYQORHGK 294
 Db 402 VTGIIISGILEASKQAMFEHVASKMADVIAEWKRGK 438

RESULT 8

ID MYSC CHICK STANDARD; PRT; 1102 AA.

AC P29616; (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Myosin heavy chain, cardiac muscle isoform (Fragment).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1] SEQUENCE OF 65-1102 FROM N.A.
 RP STRAIN=Broiler; TISSUE=Heart;
 EX MEDLINE=92130260; PubMed=1774788;
 RA Stewart A.F.R., Camoretti-Mercado B., Perlman D., Gupta M.,
 RA Jakovcic S., Zak R.;
 RT "Structural and phylogenetic analysis of the chicken ventricular
 RT myosin heavy chain rod.";
 RL J. Mol. Evol. 33:357-366(1991).
 RN [2]
 RP SEQUENCE OF 1-259.
 RC TISSUE=Heart;
 RX MEDLINE=93039740; PubMed=1418675;
 RA Watanabe B.;
 RT "Amino-acid sequence of the short subfragment-2 in adult chicken
 RT cardiac muscle myosin.";
 RL Biol. Chem. Hoppe-Seyler 373:1045-1054(1992).
 CC -|- FUNCTION: MUSCLE CONTRACTION.
 CC -|- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -|- TISSUE SPECIFICITY: VENTRICULAR MUSCLE AND TRACES LEVELS IN THE
 CC ATRIUM. ALSO TRANSIENT EXPRESSION IN SKELETAL MUSCLE DURING FETAL
 CC DEVELOPMENT AND REGENERATION FOLLOWING FREEZE INJURY.
 CC -|- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -|- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -|- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
 CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
 CC CONSERVED.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; X59552; CAA42130.1; -.
 DR HSP; P03437; 1HTM.
 DR InterPro; IPR002928; Myosin tail.
 DR Pfam; PF01576; Myosin tail.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Multigene family.
 FT NON_TER 1
 FT DOMAIN <1 1102 COILED COIL (POTENTIAL).
 FT CONFLICT 65 65 A -> T (IN REF. 1).
 FT CONFLICT 92 92 V -> A (IN REF. 1).
 FT CONFLICT 113 114 CS -> FAL (IN REF. 1).
 FT CONFLICT 135 135 H -> Q (IN REF. 1).
 FT CONFLICT 185 185 T -> R (IN REF. 1).
 SQ SEQUENCE 1102 AA; 128008 MW; 2293668D049825DC CRC64;
 Query Match 7.4%; Score 112; DB 1; Length 1102;
 Best Local Similarity 19.9%; Pred. No. 5.3;
 Matches 69; Conservative 67; Mismatches 143; Indels 68; Gaps 11;
 Qy 1 MTSIFAEQTVVVKSAIETADGALDLYNK--VLDQVIPWKT-FDETIKELSRFKQEYSOE 57
 Db 432 MTRLMNDLTQTKLQSENGEFVRQLEEKESISQLSRGKTSFTQOIEELRQLEETKS 491
 Qy 58 ASVLVGDIXVLLMD-----SODKYFEATQTVVWCVVTVLLSAYILLFDEYNK 107
 Db 492 KNALAHAIQAARHDCDLLEQYEEQEAQAEQLQALSQNAEVAQWTKYETDAQRTTE 551
 Qy 108 KSAQKDIIRI-----LDDGVKLNKAEAKSLTSSQSFNNASGKLAL 151
 Db 552 LEDAKKLLARLQEAEEAEAEANAKGSLKAKHRLQNEQEDMIDLEKANGAAA--SL 608
 Qy 152 D-----SOLTDFSEKSYFQSOVDRIKREAVAGAAAGVAGPFGILISYIAAGVIEG 205
 Db 609 DKQGFQDKIINDWKQKYEQAELASQKE-----ARSLSTELFKKNAYETLDHLB- 662
 Qy 206 KLIPELNNRLKTVQNFPTSLSATVQKQKID-----DAKKLATEIAAIGBI 253
 Db 663 -----TLKRENKQLQEBISDLTNQISEGNKLNLEIKVKQVEKSEKSEVGLALEEAGLE 718
 Qy 254 KTEETTRFVVDYDDLMLSLKGAAKMINTCNEYCQ---RHGKTL 297
 Db 719 HEESKTLRF-----QLESLQKADFERKLAEDKEMQNRNQRTI 760

RESULT 9

ID TPM MYTED STANDARD; PRT; 284 AA.
 AC Q25457;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Tropomyosin.
 OS Mytilus edulis (Blue mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
 OC Mytiloidea; Mytilidae; Mytilus.
 OX NCBI_TaxID=6550;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Anterior byssus retractor muscle;
 RA Nyitray L., Yang W., Szent-Gyorgyi A.G.;
 RA Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -|- SUBUNIT: Homodimer (By similarity).
 CC -|- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
 CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
 CC -|- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
 CC -----
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

```

KW Muscle protein; Coiled coil; Repeat; Multigene family.
SQ DOMAIN 1 284 COILED COIL.
FT SEQUENCE 284 AA; 32770 MW; 5468021A8FEED62 CRC64;

Query Match 7.3%; Score 110.5; DS 1; Length 284;
Best Local Similarity 20.6%; Pred. No. 1.2;
Matches 57; Conservative 44; Mismatches 99; Indels 77; Gaps 8;

QY 14 KSAETADGALDLYNKYLQDVIPWKVFDTEIKLSRFKQEYSQEASVLVGDIKVLLMSQ 73
Db 48 KSIQTENDLNTQTQLDQVQKYETETQKIAHEQEIQSLTRKISMLEDI----MKSE 103
QY 74 DKYEATQTVYEW-----CG---VVTQLLSAYILLFDEYNEKKASQAQ 112
Db 104 ERYTTAAASKLEASKADESERNRKVLNLCNGNDERIDQL-----EKQLTEA 151
QY 113 KDILRIIDGCVKYLNAQKSLTSSQSFNNAASKLLDLSOLDTDFSE-----KS 163
Db 152 K---WIAEADKKVEAARKLATIEVDLEAEARLEAAEKVI-DLEQLTVVGANIKT 206
QY 164 SYFQSQVDRIKEAYAGAAAGIVAGPPLGIISYIAAGVIEGKLIPELNNRLKTVQNFTT 223
Db 207 LQVNDQASQSDSEVE-----ETIRDLTNRKDAENRAT 240
QY 224 SLSATVQKANDIDAANKLKLATETAAIGETKETT 260
Db 241 EAERTVSKLQEVDRLEDELLTKEKYKKAISDELDA 277

RESULT 11
ID_SVR_SVNY3
ID_SVR_SVNY3 STANDARD; PRT; 584 AA.
AC Q55486;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
GN ARG3 OR SLL0502.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugliura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -I- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -I- SUBUNIT: Monomer (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D64006; BAA10833.1; -.
CC PIR; S75986; S75986.
CC HSP; Q05506; 1B82.
CC HMAP; NP_00123; -.
CC InterPro; IPR001278; Arg_tRNA-synt_lc.
CC InterPro; IPR005148; N.
CC InterPro; IPR001412; tRNA-synt I.
CC Pfam; PFC03485; N-Arg; 1.
CC Pfam; PFC0750; tRNA-synt 1d; 1.

```



```
SQ SEQUENCE 1938 AA; 223255 MW; 387399C8F63A4CF4 CRC64;
Query Match 7.3%; Score 110; DB 1; Length 1938;
Best Local Similarity 18.9%; Pred.No.14;
Matches 55; Conservative 60; Mismatches 114; Indels 62; Gaps 9;
QY 5 FAEQTEVVK-----SAIETADGALDLYNKVDQVWPWKTFDETIKLSRPFK 51
DB 1315 FSSQVEAKAABDELHERQEFHAACKNLEHLDCHELLEQINGK--DDIQQLSRIN 1372
QY 52 QEYSQ-----EASVLVG--DIKVLMLDSODKYFEATQTVYVWGVVTLQSLAVILLEDEY 104
DB 1373 SEISQWAKYVEGSLVSELELKKQKNRVMDLOEALSAQNKVISEKAKGLLAET 1432
QY 105 NEKKAQAQDILI-----RILDGVKKLNEAKSLTSSQNNASGKLLALD 152
DB 1433 EDARSVDRLITVIALEKKQKRAFQKIVDDWKEKVDIQKEIDATDRGNTSTEVFKLR 1492
QY 153 SOLTNDPSEKSSVFSQVDRIKREAYAGAAAGIVAGPFGILLISYIAAGVIE-----GKL 207
DB 1493 SSMND-----LSQIETLRE-----NKIFSQEL-RDINEQITCGGRT 1529
QY 208 IPELNRLKTVQNFPTSLSNVQKANKDIDAAKLKLAETIAAIGEIKTETE 258
DB 1530 YQEVHKSVRRLQEKDELQHALDEABAALAEESKVLRLQIEVQOQIRSEIE 1580

RESULT 13
GCG_HUMAN STANDARD; PRT; 2230 AA.
AC Q13439; Q13654; Q14436;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa
DE golgin) (Golgin-245) (72.1 protein).
GN GOLG44.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=96215236; PubMed=8628529;
RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT "Molecular characterization of trans-Golgi p230: a human peripheral
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif.";
RL J. Biol. Chem. 271:8328-8337(1996)
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzel M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
RT protein containing a granin signature.";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balague C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC -!- FUNCTION: May play a role in vesicular transport from the trans-
CC Golgi.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
CC GOLGI MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
```

```
QY 204 EGKLIPELNNRLKTVQNFSTLSATVKQANKDIDAAL--KLAETIAAIGEI---KTET 257
Db 1167 DKRKVSELTSLKUTDEBQSLKSHKSNKSLDSEFKKLSEELAQLDICKKTE- 1225
QY 258 ETTREYVDYDDMLSLKGAARKMIN 283
Db 1226 -----ALLEAKTNELIN 1237

RESULT 14
ID LKAA PASHA STANDARD; PRT; 955 AA.
AC P55117;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leukotoxin from serotype T10.
GN LKTA.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Mannheimia.
OX NCBI_TaxID=75985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype T10;
RX MEDLINE=96425875; PubMed=8828217;
RA Lainsion F.A., Murray J., Davies R.C., Donachie W.,
RT "Characterization of epitopes involved in the neutralization of
RT Pasteurella haemolytica serotype A1 leukotoxin.";
RL Microbiology 142:2493-2507(1996).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: The Gly-rich region is probably involved in binding
CC calcium, which is required for target cell-binding or cytolytic
CC activity.
CC -!- DOMAIN: The three transmembrane domains are believed to be
CC involved in pore formation by the cytotoxin (BY SIMILARITY).
CC -!- PTM: Palmitoylated by lktC. The toxin only becomes active when
CC MODIFIED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z26247; CAA81206.1; -.
CC FIR; S37145; A35254.
CC InterPro; IPR001343; Hemlysn_Ca_bind.
CC InterPro; IPR003995; RTXa.
CC Pfam; PF00353; hemolysinCbind; 5.
CC Pfam; PF02382; RTX; 1.
CC PRINTS; PR00313; CABNDNGRPT.
CC PROSITE; PRO1488; RTXTOXINA.
CC PROSITE; PS00330; HEMOLYSIN_CALCIUM; 4.
CC Hemolysis; Toxin; Cytolysis; Repeat; Calcium; Transmembrane;
KW Lipoprotein; Palmitate.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 361 381 POTENTIAL.
FT TRANSMEM 383 403 POTENTIAL.
FT DOMAIN 736 786 6 X REPEATS, GLY-RICH.
FT REPEAT 736 741 1.
FT REPEAT 745 750 2.
FT REPEAT 754 759 3.
FT REPEAT 763 768 4.
FT REPEAT 772 777 5.
FT REPEAT 781 786 6.
SQ SEQUENCE 955 AA; 102187 MW; B60F2DB8168EBCAF CRC64;
```

```
Query Match 7.2%; Score 109; DB 1; Length 955;
Best Local Similarity 18.3%; Pred. NO. 7;
Matches 68; Conservative 63; Mismatches 139; Indels 102; Gaps 11;
QY 6 ABQTVVWKSATETADGALDLYNKYLDQVTPKTFDETIKELSRPKQEQYSQASVLVG-- 63
Db 89 AQTSLGTIQNVLGITRGIVLSAPQLDKLLQKNKYGQALGSSSESTAQNFQSKTVLSGVQ 148
QY 64 -----DIKVLMDSDKY-----FEATQTVYEWGVVTVQLLSAYILLFDYNEK 107
Db 149 GNSRTVLAGMDLDEALQNESDQTLTAKAGLELTNSLIENIANSVOTLDAFSEQISFGSK 208
QY 108 -----KASAKDILIRILDGVKKLNEAKOKSLITSSQSFNNASG 146
Db 209 LQNVKGLGALGDKXKNIGGLDRAGLGLDVKSLLSCATLVLAKDKASTAK--VG 263
QY 147 KLLALDSQLTNDFSEK-SSYFOSQVDRIRKEAYAGAAAGIVA-GPGLIISYSIAAGV-- 202
Db 264 AGPELANQVVGNIITKAVSSVILAQ-----RVAAGLSSTGPVAALIASVVAISP 313
QY 203 IEGKLIPELNNRLKTVQNF-----TSLSATVKQANKDIDAALKLATEAAI-- 250
Db 314 LSPAGIADKFDRAKSLDENTAEKPKLGYEGDGLLAEYQHGTTIDASVTAINTALLAAG 373
QY 251 -----GEIKTETETTRFYVDYDDMLSLKGAARKMINTCNE 287
Db 374 GVSAAGSVAASPIALLVSGITGVISTLIQYSK-----QAMFEHVANKIHKNIVE 424
QY 288 YQQRHGHKKTLFE 299
Db 425 WEKNGGKNYFE 436
```

RESULT 15

```
KL61_DROME STANDARD; PRT; 1066 AA.
AC P4683; Q8TU06; Q9W018;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bipolar kinesin KRP-130 (Kinesin-like protein klp61F).
GN KLP61F OR KLP2 OR CG9191.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94043448; PubMed=8227131;
RA Heck M.M.S., Pereira A., Pesavento P.A., Yannoni Y., Spradling A.C.,
RA Goldstein L.S.B.;
RT "The kinesin-like protein KLP61F is essential for mitosis in
RT Drosophila.";
RL J. Cell Biol. 123:665-679(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej B.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.P., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA Balleson R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
```

de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann M.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodsk A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpman G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrafi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassaman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RA [3]
RA REVISIONS.
RC STRAIN=Berkley; PubMed=12337572;
RX MEDLINE=22426069; PubMed=12337572;
RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.B.;
RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RA [4]
RA SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12337569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celisner S.E.;
RT "A *Drosophila* full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RA [5]
RA SEQUENCE OF 228-357 FROM N.A.
RC STRAIN=DP CN BW;
RX MEDLINE=92020874; PubMed=1924306;
RA Stewart R.J., Pesavento P.A., Woerpel D.N., Goldstein L.S.B.;
RT "Identification and partial characterization of six members of the
RL kinesin superfamily in *Drosophila*.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8470-8474(1991).
RA [6]
RA PARTIAL SEQUENCE, AND IDENTIFICATION AS KRP-130.
RX MEDLINE=97078747; PubMed=8918872;
RA Kashina A.S., Scholey J.M., Leszyk J.D., Saxton W.M.;
RT "An essential bipolar mitotic motor.";
RL Nature 394:225-225(1996).
CC -1- FUNCTION: IMPORTANT ROLE IN MITOTIC DIVIDING CELLS. MICROTUBULE
CC MOTOR REQUIRED FOR SPINDLE BODY SEPARATION. SLOW PLUS-END DIRECTED
CC MICROTUBULE MOTOR CAPABLE OF CROSS-LINKING AND SLIDING APART
CC ANTIPARALLEL MICROTUBULES, THEREBY PUSHING APART THE ASSOCIATED
CC SPINDLE POLES DURING SPINDLE ASSEMBLY AND FUNCTION.
CC -1- SUBUNIT: Homotetramer.
CC -1- DEVELOPMENTAL STAGE: SPECIFICALLY EXPRESSED IN PROLIFERATING
CC TISSUES DURING EMBRYONIC AND LARVAL DEVELOPMENT.
CC -1- PM: PHOSPHORYLATION DURING MITOSIS AT THR-933 CONTROLS THE
CC ASSOCIATION OF Klp61f WITH THE SPINDLE APPARATUS (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIRC
SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U01842; AAA03718.1; -
CC EMBL: AE003471; AAF47458.2; -
CC EMBL: AI069442; AAL39587.1; -
CC EMBL: M74428; AAA28655.1; -
CC PIR: A48669; A48669.
CC FLYBASE: P17119; 3KAR.
CC FLYBASE: FBGN004378; Klp61f.
CC GO: GO:0005737; C:cytoplasm; IDA.
CC GO: GO:0005871; C:kinesin complex; IDA.
CC GO: GO:0003774; F:motor activity; IDA.
CC GO: GO:0007100; P:centrosome separation; IGI.
CC InterPro: IPR001752; kinesin_motor.
CC Pfam: PF00223; kinesin; 1.
CC PRINTS: PR00380; KINESINHEAVY.
CC SMART: SMC0129; KISC; 1.
CC PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE: PS50067; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW Mitosis; Cell cycle; Phosphorylation.
FT DOMAIN 17 361 KINESIN-MOTOR (BY SIMILARITY).
FT DOVAIN 362 462 COILED COIL (POTENTIAL).
FT DOVAIN 540 569 COILED COIL (POTENTIAL).
FT DOVAIN 639 738 COILED COIL (POTENTIAL).
FT DOVAIN 808 875 COILED COIL (POTENTIAL).
FT DOVAIN 889 918 COILED COIL (POTENTIAL).
FT NP_BIND 103 110 ATP (POTENTIAL).
FT MOD_RES 933 933 PHOSPHORYLATION (BY CDC2) (BY
SIMILARITY).
FT VARIANT 595 595 M -> V.
FT VARIANT 869 869 R -> K.
FT VARIANT 904 904 H -> Q.
FT CONFLICT 962 962 L -> Q (IN REF. 1).
FT CONFLICT 983 983 V -> D (IN REF. 1).
SQ SEQUENCE 1066 AA; 121163 MW; 363647366EE0721F CRC64;
Query Match 7.2%; Score 108.5; DB 1; Length 1066;
Best Local Similarity 21.6%; Pred. No. 8.7;
Matches 62; Conservative
QY 24 LDINKYLDQVFPKTFDETIKLSRFQYSEASVLVDIKVLLMDSQDYFEATQTV 83
DB 159 LELYNEELCDLL---STDDTK--IRIFDDSTKGSVIOGLEIPVHSKDDVYKLEKG 213
QY 84 YEWGVVTVQLLSAYILLFDEYNEKKAQAQDIILIRILDGK-----KLAENACKSLTSS 138
DB 214 KERKATATLMA-----QSRSHVPSIVVHIRENGIEGEDMLKIKLNLVDLAGS 265
QY 139 QSFNNAGK-----LLALDSQLTN--DFSEKSSYFQSVQDRIKRAYAGAA 182
DB 266 ENVSAGNEKGIKRVETVNVNQSLTLGRVITALVDRAHPVYRESKTLTLOESLGRT 325
QY 183 -----AGIVAGFFGL-----IISYSIAAGVIEGKLIPELNNRL--KTVQNFFTSLSATVK 230
DB 326 KTSIATISPKKIDETLTSTLEYAHRAKNQNK--PEVQKLTAKTVLKEYTE---EID 380
QY 231 QANKIDAAKUKLATEIA--AIGIKETETTTREYVDYDMLSLK 275
DB 381 KLRDLMAARDKNGIYLAETTYGTEITLKESQRELNEKMLLKALK 427
RESULT 16
MLFL_YEAST

ID MLP1_YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=9247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.V., Botstein D.;
RL "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RL "The complete sequence of a 15,820 bp segment of Saccharomyces
RL cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RL new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REP.2 MISQUOTES THE GENE NAME AS "MLP1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L01992; AAA34783.1; -
DR EMBL; Y73541; CAAS1948.1; -
DR EMBL; Z28320; CAA82174.1; -
DR PIR; S38173; S38173.
DR SGD; S0001803; MLP1.
DR GO; GO:0005635; C:nuclear membrane; IDA.
DR GO; GO:0006554; C:nucleoplasm; IDA.
DR GO; GO:0006006; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 7.1%; Score 107.5; DB 1; Length 1875;
Best Local Similarity 19.9%; Pred. No. 20;
Matches 64; Conservative 49; Mismatches 127; Indels 81; Gaps 12;

QY 7 EQTEVVKSALETADGALDLYNKYLDQVPMKTDTEI-----KELSRFKQYEQAES 59
DB EDKISLLKEQMFNLNNELDLQKGWEK--EKADFKRISILQNNKEVEAVKSEYSEKLS 1023
QY 60 VLVGDIKVLMDSDQKYFEATQYVWCVVQTLLSAVILLFDYNEKKAOKDI----- 115
DB KIQND-----LDQQTIVANTQNNYE-----QELQKHADVSKTISLR 1061
QY 116 -LIRILDGVKLN-----EAKSLTSSQSFNNASGKLLA---LDSQLTNDPSEKSYFQ 167
DB EQLHTYKGVKTLNLSRDLENALKENKESWSQKESLLEQLDLSNRIEDLSQNKLLY 1121
QY 168 SQVDRIKAEVAGAGIVGFFGLIISYSIAAGVIEGKLIPELNLKTVQNPFFLSA 227

Db 1122 DOI-----QIYTAADKEVNNSTNG-----PGLNNLILIRRRDILDT 1159
QY 228 TVQANKDIDAALKLATEIAAIGEIKETETTR-----FYVDYDDM-----ISLL 274
Db 1160 KVTVAERDAKMLRKQISLMDVELQDARTKLDNSRVEKENHSSIIQQHDDIMEKLNQLL 1219
QY 275 KGAACKMINTCNEYCQRHGKK 295
Db 1220 R---ESNITLNELENNNNKK 1237

RESULT 17
SSP5_STRGN
ID SSP5_STRGN STANDARD; PRT; 1500 AA.
AC P169E2; Q54194;
DT 01-AUG-1990 (Rel. 15, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Agglutinin receptor precursor (SSP-5).
GN SSP5 OR SSPB.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE=90236997; PubMed=2185241;
RA Demuth D.R., Golub E.E., Malamud D.;
RL "Streptococcal-host interactions. Structural and functional analysis
RL of a Streptococcus sanguis receptor for a human salivary
RL glycoprotein.";
RL J. Biol. Chem. 265:7120-7126(1990).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE=96310377; PubMed=8733238;
RA Demuth D.R., Duan Y., Brooks W., Holmes A.R., McNab R.,
RA Jenkinson H.F.;
RL "Tandem genes encode cell-surface polypeptides SspA and SspB which
RL mediate adhesion of the oral bacterium Streptococcus gordonii to
RL human and bacterial receptors.";
RL Mol. Microbiol. 20:403-413(1996).
CC -!- FUNCTION: MAY BIND SALIVARY ACID RESIDUES OF SALIVARY AGGLUTININ
CC (SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH
CC ITS RECEPTOR IN VARIOUS ORAL STREPTOCOCCI MODULATE BACTERIAL
CC COLONIZATION OF ORAL TISSUE AND IS ASSOCIATED WITH REDUCED LEVELS
CC OF DENTAL CARIES.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -!- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U40026; AAC44100.1; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGSFAM6; TIGS01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Peptidoglycan-anchor; Calcium-binding; Repeat; Signal.
FT SIGNAL 1 38
FT CHAIN 39 1470 AGGLUTININ RECEPTOR.
FT PROPEP 1471 1500 REMOVED BY SORTASE (POTENTIAL).
FT DOMAIN 164 470 4 X APPROXIMATE TANDEM REPEATS, HRI.
FT REPEAT 164 241 1.
FT REPEAT 242 323 2.

FT REPEAT 324 405 3.
FT REPEAT 406 470 4.
FT DOMAIN 771 887 3 X APPROXIMATE TANDEM REPEATS, PR1.
FT DOMAIN 1414 1436 PRO-RICH (PR2).
FT CA_BIND 220 235 POTENTIAL.
FT CA_BIND 301 316 POTENTIAL.
FT CA_BIND 931 950 POTENTIAL.
FT CA_BIND 1300 1315 POTENTIAL.
FT SITE 1467 1471 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1470 1470 AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1500 AA; 164552 MW; DCF190E7D44D899F CRC64;

Query Match 7.0%; Score 106.5; DB 1; Length 1500;
Best Local Similarity 23.5%; Pred. No. 18;
Matches 78; Conservative 42; Mismatches 97; Indels 115; Gaps 18;

QY 15 SALETADGALDLYNKYLQDVIPWKTDFETIKELSRFKOE---YSGEASVLYGDIKVLMD 71
DB 123 TATDRAQKDEIKSYAKQAEBIKTTTEA-----YKVEAAHQAEPTKINAENKA--- 172
QY 72 SODKYFEATQTVYVWGVVTVQLLSAYILLFDEYNKKSAAQDILIRILDGVKKLNEAQ 131
DB 173 ADDKYQKOLKSHOE---EVEKINTANATAKAEYEAKLAQYQKDLAT-----VKKANE-- 221
QY 132 KSLITSSQSFNNASGKLLALDSQNDSEKSSYFQSOVDLR-----KEAYAGAAAGI 185
DB 222 -----DSQ---QDYQNKUSAYQTELARVQKNAEAKAEYKAVKEN 259
QY 186 VAGPFGLIISYISIAAGVIEGKLIPELNNRLK-----TVONFFTSLSATVKQAKDIDA-A 239
DB 260 TAKNEAL-----KVENEAIKORNETAKATYEAMKQYEAADLAA-IKKANEDNDADY 309
QY 240 KKLIA---TEIAIGE 265
DB 310 QAKLAAYQTELARVQKNAEAKAEYKAVKENAKNTAKNTAIQAEINEAIKORNETAKATYDAA 369
QY 266 ---YD-DLMLSLKGAACKMINTCNE--YQOR 291
DB 370 VKYKYEADL-----AAVQKQATNEADYQAK 394

RESULT 18
LKA3_PASHA STANDARD; PRT; 953 AA.
AC P55116;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leukotoxin from serotype T3.
GN LKTA.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Mannheimia.
OX NCBI_TaxID=75985;
RN SEQUENCE FROM N.A.
RP STRAIN=serotype T3;
RC MEDLINE=94041617; PubMed=8225575;
RA Burrows L.L., Lo R.Y.C., Olah-Winfield E.;
RT "Molecular analysis of the leukotoxin determinants from Pasteurella
haemolytica serotypes 1 to 16."
RL Infect. Immun. 61:5001-5007(1993).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
cell membranes and cause cell rupture by mechanisms not clearly
defined.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: The Gly-rich region is probably involved in binding
calcium, which is required for target cell-binding or cytolytic
activity.
CC -!- DOMAIN: The three transmembrane domains are believed to be
involved in pore formation by the cytotoxin (BY SIMILARITY).
CC -!- PTM: Palmitoylated by IkcC. The toxin only becomes active when
modified (By similarity).

CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR ENBL: U01216; AAB36691.1; --
DR InterPro: IPR001343; Hemlysn_Ca_bind.
DR InterPro: IPR003995; KtxA.
DR Pfam: PF00353; hemolysin_cbind; 5.
DR Pfam: PF02382; RTX; 1.
DR PRINTS: PR00313; CABNDNGRPT.
DR PRINTS: PR01488; RTXTOXINA.
DR PROSITE: PS00330; HEMOLYSIN_CALCIIUM; 2.
DR Hemolysis; Toxin; Cytolysis; Repeat; Calcium; Transmembrane;
Lipoprotein; Palmitate.
FT TRANSMEM 229 249 POTENTIAL.
FT TRANSMEM 297 318 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT DOMAIN 734 784 6 X REPEATS, GLY-RICH.
FT REPEAT 734 739 1.
FT REPEAT 743 748 2.
FT REPEAT 752 757 3.
FT REPEAT 761 766 4.
FT REPEAT 770 775 5.
FT REPEAT 779 784 6.
SQ SEQUENCE 953 AA; 101948 MW; FDBDCB2FDC85PDF2 CRC64;

Query Match 7.0%; Score 106; DB 1; Length 953;
Best Local Similarity 18.8%; Pred. No. 11;
Matches 68; Conservative 61; Mismatches 152; Indels 80; Gaps 10;

QY 6 AEQTEVVKSAIETADGALDLYNKYLQDVIPWKTDFETIKELSRFKQEVYSGEASVLYG-- 63
DB 87 AQTSLGTIQNVLGITRGIIVLSAPQLDKLLQKNKVGQALGSSBSIAQNFQSKTVLSGVQ 146
QY 64 -----DIKVLMDSDQKY-----FEATQTVYVWGVVTVQLLSAYILLFDEYNEK 107
DB 147 SILGSVLAMGMDLEALQNESDQTLKAGLELTNSLIENIANSVQTLDAFSEIGISQFGK 206
QY 108 KASAAQKDLIRILDGVKK---LNEAQKSLITSSQSFNNASGKLALD----- 152
DB 207 LQNVKG---LGALGDKLKNIGGLDKAGLGLHVISGLLSGATAALVLADKDASTAKKVAG 263
QY 153 ----SQTNDPSEK--SSYFQSOVDRIKEAYAGAAAGIVAGPFGLIISYISIAAGVIE--- 204
DB 264 FELANQVGNITKAVSSYLLAQVAARLS--TGVAALTAFTVALAISPLSFAGIADKFD 322
QY 205 -GKLIPELNNRLKTVONFFTSLSATVKQAKDIDAALKLATEIAAI----- 250
DB 323 RAKSLENYAERFKLGYEGDLSLAELYQHGTGTIDASVTAINATAAIAAGVSAAGSVV 382
QY 251 -----GEIKTEFTTFYDDYDDMLSLKGAACKMINTCNEVQORHGKKTLF 298
DB 383 ASPALLVSGITGVISTILQYSK-----QAMFEHVANKLNKIVKEKNGKGNVF 433
QY 299 E 299
DB 434 E 434

RESULT 19
ESG2_TRYBB STANDARD; PRT; 329 AA.
ID ESG2_TRYBB
AC P04478;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE VSG expression site-associated protein 221a precursor (ESAG protein).

OS Trypanosoma brucei.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85254917; PubMed=2861910;
 RA Cully D.F., Ip H.S., Cross G.A.M.;
 RT "Coordinate transcription of variant surface glycoprotein genes and
 RL Cell 42:173-182(1985).
 CC -!- FUNCTION: NOT KNOWN BUT MAY BE RELATED TO ACTIVATION OF THE
 CC VARIANT SURFACE GLYCOPROTEIN GENES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M11452; AAA30191.1; -;
 DR PIR; A03395; VMUT21.
 DR InterPro; IPR004922; ESAG.
 DR Pfam; PF03238; ESAG1; 1.
 KW SIGNAL.
 FT CHAIN 1 23
 FT CHAIN 24 329 VSG EXPRESSION SITE-ASSOCIATED PROTEIN
 FT 221A.
 FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 329 AA; 36603 MW; 4D19F59477D9CEB8 CRC64;
 Query Match 6.9%; Score 104; DB 1; Length 329;
 Best Local Similarity 21.9%; Pred. No. 4.1;
 Matches 46; Conservative 34; Mismatches 64; Indels 56; Gaps 9;
 QY 92 QLLSAYLLFD-----EYNEKASQAKDILIRLDGKKL-NEAQS 133
 DB 4 EIVELWLLFSVTCVDALQAGDCTRVADHKEHAPVTEAVCYLCLSLDKLYSEGEKK 63
 QY 134 LITSQSNNASGKLLALDLSLTNDPSEKSY-----FOSQVDRIRKEAYAGAAAGV 186
 DB 64 LLVTEEVYANAS---LLILD-DMEGRAGESSYLSVIRGWEEQDRLEK----- 108
 QY 187 AGPGLIISYSIAAGVI---EGKLIPELNNRLKTQV-----NFTTSLSATVQA 232
 DB 109 -----LISYGNKQGNLVAKAGGLFAALDSLSKEVRKEIPGALIKTKNYTSVAEIVRTV 162
 QY 233 NKDIDAANKL-----ATEIAATGEIKTE 256
 DB 163 WEDVGEILLWKEATKCGSQKVGEGEIQTE 192
 RESULT 20
 ID WAPA_BACSU STANDARD; PRT; 2334 AA.
 AC Q07833.
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Wall-associated protein precursor.
 GN WAPA OR N17G.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=168;
 RX MEDLINE=93302506; PubMed=8316082;
 RA Foster S.J.;
 RT "Molecular analysis of three major wall-associated proteins of

RT Bacillus subtilis 168: evidence for processing of the product of a
 RT gene encoding a 258 kDa precursor two-domain ligand-binding
 RT protein.";
 RL Mol. Microbiol. 8:299-310(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 RT genome containing the hut and wpa loci.";
 RL Microbiology 141:337-343(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=168 / BGSC1A1;
 RX MEDLINE=97124196; PubMed=8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 RA Miwa Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the lic and cel loci, and creation of a 177 kb contig
 RT covering the gnt-sacXY region.";
 RL Microbiology 142:3113-3123(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klueter P., Koningstein G., Krogh S., Kumano M.,
 RA Kobayashi Y., Koetter P., Lardinis S., Lauber J., Lazarevic V.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seter S.J., Sertor P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Totsato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 CC MOTILITY, SECRETION OR DIFFERENTIATION
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 CC INTO THE MEDIUM.
 CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 CC MOTIF REPEATED 31 TIMES.
 CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 CC SIMILARITY TO THE REPEAT IN E. COLI RHS GROUP OF PROTEINS (RHS-A-D).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC CC EMBL; L05634; BAA22883.1; -
CC CC EMBL; D31856; BAA06656.1; -
CC CC EMBL; D29985; BAA0620.1; -
CC CC EMBL; D83026; BAA11683.1; -
CC CC EMBL; Z99124; CAB15959.1; -
CC CC F1R; S32920; S32920.
CC CC Subtilisin; BG10797; wapa.
CC CC InterPro; IPR003305; CEM_Cenc.
CC CC Pfam; PF02018; CEM_4_9_1.
CC CC TIGRFAMs; TIGR01643; YD_repeat_2x; 17.
KW Cell wall; Repeat; Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 504 605 1-1.
FT REPEAT 636 736 1-2.
FT REPEAT 769 869 1-3.
FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
FT REPEAT 1021 1040 2-1.
FT REPEAT 1042 1061 2-2.
FT REPEAT 1063 1082 2-3.
FT REPEAT 1083 1102 2-4.
FT REPEAT 1103 1128 2-5.
FT REPEAT 1129 1148 2-6.
FT REPEAT 1150 1169 2-7.
FT REPEAT 1174 1193 2-8.
FT REPEAT 1199 1218 2-9.
FT REPEAT 1219 1238 2-10.
FT REPEAT 1646 1665 2-11.
FT REPEAT 1667 1686 2-12.
FT REPEAT 1690 1709 2-13.
FT REPEAT 1711 1730 2-14.
FT REPEAT 1732 1751 2-15.
FT REPEAT 1753 1772 2-16.
FT REPEAT 1795 1814 2-17.
FT REPEAT 1820 1839 2-18.
FT REPEAT 1840 1859 2-19.
FT REPEAT 1861 1880 2-20.
FT REPEAT 1887 1906 2-21.
FT REPEAT 1908 1927 2-22.
FT REPEAT 1929 1948 2-23.
FT REPEAT 1969 1982 2-24 (APPROXIMATE).
FT REPEAT 1983 2002 2-25.
FT REPEAT 2008 2027 2-26.
FT REPEAT 2028 2047 2-27.
FT REPEAT 2051 2070 2-28.
FT REPEAT 2071 2090 2-29.
FT REPEAT 2093 2112 2-30.
FT REPEAT 2120 2139 2-31.
CC CC SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;

Query Match 6.9%; Score 104; DB 1; Length 2334;
Best Local Similarity 25.8%; Pred. No. 46; Mismatches 85; Indels 50; Gaps 13;
Matches 60; Conservative 39;

QY 78 EATQTV-YEVCVVYVQLLSAYILLDFEYNEKK-ASAQKDLIRILD---DGVKLNKAEAK 132
DB 1203 KATESQYDKDGNVTGKDAYGTVEYKNDVTKMKTGNTGNTDIAVDGLDAVSETDQ 1262
QY 133 SLTSS---QSNN--ASGLIALDSQLTNDFS---EKSSY-FQSQVDRIKEAYAGAA 182
DB 1263 SGKSSGAADVTKGNQIQSKDLSASTNLIKGSFPAQSKGNLTKASKDR-FKISVIADK 1321
QY 183 AGIVAGPFGI-IISYSIAAGVIEGKLIPELNNRLKTVQVFTSLSATVKAQNKDIAAKL 241
DB 1322 SGVLGSKALEVLQSSTAGTDHG-----YSSATQTV-----EL 1355
QY 242 KLATEIAAIGEIKETETTRTFYVDYDDLMLSLKGAAGKQVINTC-NEYQORHGK 294
```

```
DB 1356 EPNTTYTSLGKIKTDLAKSRAYFNID-----LRDKDQRIQWIHNEYSALAGK 1403

RESULT 21
RECN_AQUAE STANDARD; PRT; 520 AA.
AC O66834;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA repair protein recN (Recombination protein N).
GN RECN OR AQ_561.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
NCBI_TaxID=63363;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=VF5;
RC MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead W.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358 (1998).
CC -!- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED
CC -!- DNA (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RECN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; A800695; AAC06789.1; -
DR PIR; F70350; F70350.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR004604; RecN
DR InterPro; IPR003405; SMC_C.
DR Pfam; PF02483; SMC_C; 1.
DR TIGRFAMs; TIGR00634; recN; 1.
KW DNA repair; ATP-binding; Complete proteome.
FT NP BIND 29 36 ATP (POTENTIAL).
SQ SEQUENCE 520 AA; 60439 MW; A9708562ACBE901E CRC64;

Query Match 6.8%; Score 103.5; DB 1; Length 520;
Best Local Similarity 19.4%; Pred. No. 7.7; Mismatches 95; Indels 71; Gaps 13;
Matches 57; Conservative 71;

QY 10 VEVKSAIETADGALDLYNKYLDQVLPWKTFDE---TIKLSRFKQSYQ---EASVL- 61
DB 135 LEKYNLSRKKEQELFEFLRKEELIQQDYLEFRVREVEEIGISSEVEELKNKANLIN 194
QY 62 -VGDIKVLLMDSQKQFEATQTVYEWGVVTVQLS---AYILLDFEYNEKKAQKDI-- 115
DB 195 NLEKVKAVGESLYKLEGENSVYEILGEIRKNLAKVESYSGKPSFLIKIANLEEVVE 254
QY 116 -----LIRILDGQVKLNKAEAKSLTSSQSFNNASGKLL---ALDSQLTN----DF 159
DB 255 LYNLSKEMPEISEEVEVNEINEKLFRIQRLKEKYKSPFELKEVEIKEELSNLSVDF 314
QY 160 SEKSSYFQSQVDRIKEAYAGAAAGVAGPFGILLIISYTAAGVIEGKLIPELN-NELKTV 218
DB 315 KEE- -LRESEVKLREE-----
QY 219 QNFTTSLSATVKAQNKDIAAKLKLATEIAAIGEIKETETTRTFYVDYDDLMLSL 272
DB 346 EDLEERTEELKELN--LERAKLV-----EIK-ESEPTYGKDKIEFLFS 388


```

RESULT 22

```
MYSN_DROME STANDARD; PRT; 2017 AA.
ID Q99323;
AC 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE MYOSIN heavy chain, non-muscle (zipper protein) (Myosin II).
GN ZIP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC TISSUE=Embryo;
RX MEDLINE=90349606; PubMed=2117279;
RA Kechum A.S., Stewart C.T., Stewart M., Kiehart D.P.;
RT "Complete sequence of the Drosophila nonmuscle myosin heavy-chain
RT transcript: conserved sequences in the myosin tail and differential
RT splicing in the 5' untranslated sequence."
RL Proc. Natl. Acad. Sci. U.S.A. 87:6316-6320(1990).
CC -!- FUNCTION: NONMUSCLE MYOSIN APPEARS TO BE RESPONSIBLE FOR
CC CELLULARIZATION. REQUIRED FOR MORPHOGENESIS AND CYTOKINESIS.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q99323-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q99323-2; Sequence=VSP 003342;
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35012; AAA28713.1; -.
CC PIR; A36014; A36014.
CC HSP; P10587; 1BR2.
CC FlyBase; FBgn0005634; zip.
CC GO; GO:0005856; C:cytoskeleton; NAS.
CC GO; GO:0005860; C:non-muscle myosin; NAS.
CC GO; GO:0030018; C:Z disc; IDA.
CC GO; GO:0007391; P:dorsal closure; IMP.
CC GO; GO:0006936; P:muscle contraction; IMP.
CC GO; GO:0045214; P:sarcomere organization; IMP.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR001609; myosin_head.
CC InterPro; IPR004009; myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC Pfam; PF00612; IQ; 1.
CC Pfam; PF00663; myosin_head; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC Pfam; PF01576; Myosin_tail; 1.
CC PRINTS; PR00193; MYOSINHEAD.
CC PRODOM; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS00096; IQ; 1.
CC Myosin; Alternative splicing; Coiled coil; Actin-binding;
KW ATP-binding; Calmodulin-binding.
KW DOMAIN 1 829 MYOSIN HEAD-LIKE.
FT DOMAIN 830 859 IQ.
FT DOMAIN 886 2017 COILED COIL (POTENTIAL).
FT NP_BIND 225 232 ATP.
FT DOMAIN 250 260 25 kDa/50 kDa JUNCTION.
```

```
FT DOMAIN 682 694 50 kDa/20 kDa JUNCTION.
FT DOMAIN 705 727 ACTIN-BINDING.
FT DOMAIN 742 758 REACTIVE SULFHYDRYL/ACTIN-BINDING.
FT DOMAIN 1303 2017 LIGHT MEROMIOSIN (LMM).
FT DOMAIN 1303 1970 ALPHA-HELICAL TAILPIECE (LMM).
FT DOMAIN 1971 2017 GLOBULAR TAILPIECE.
FT VARSPLIC 1 45 Missing (in isoform Short).
FT VARSPLIC 1 45 /FTID=VSP 003342.
SQ SEQUENCE 2017 AA; 232016 MW; 73E3CE02BA8F2528 CRC64;
Query Match 6.8%; Score 103.5; DB 1; Length 2017;
Best Local Similarity 19.9%; Pred. No. 41;
Matches 74; Conservative 53; Mismatches 112; Indels 133; Gaps 15;
QY 8 QTVEVVKGAETADGALDLYNKYL-----DOVIFPKTFDETIK 45
Db 1257 ENLRKAKTVLEKAGKTLEAENADLATELSVNSRQENDRRRKQAESQIAELQV---KLA 1313
QY 46 ELSRFKQESQASVLVDGKVLKMDSDQKFEATQTVYWGCVVTVLLSAYILLFDYIN 105
Db 1314 EIERARSELQEKCTKLQOEAEINITNQLAEELKASAAVKSASNMESQLTEAQQLLEBETR 1373
QY 106 EK-----KASAKQDI-LIRI 119
Db 1374 QKLGLSSKLRLQIESEKEALQELBEDDEAKRNYERKLAEVITQGEIKKAEEDADLAKE 1433
QY 120 LDGVKKLN-----EAQ-KSLTSSQSFNNASGKL-----LALDSQLTNDFS-EKS 163
Db 1434 LEEGKRLNKKQTEALERQVKELIAQNRLDKSKKIQSELEDATIELEAQRTKVLELEKK 1493
QY 164 SYFQSQVDRIKEAYAGAAAGVPGFGLIISYSTA-----AGVIEGKL----- 207
Db 1494 ----QKNFKILAEKVA-----ISEQIAQRDTAEREAREKEKTVLSVREL 1536
QY 208 -----IPELNRLKTVQNFTLSATYKQANKOI-DAAKKLAETIAAIGETETETT 260
Db 1537 DEAFKIDLEENKRTKLQNELDLANTQOTADKNVHELEKAKRALE-SQLAEKQNEEL 1595
QY 261 RFVVDYDDLMLS 272
Db 1596 E-----DDLQLT 1602
RESULT 23
Y328 MYCGE STANDARD; PRT; 756 AA.
AC Q494I9; Q49308; Q49320;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG328.
GN MG328.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.W., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 19-113 AND 155-272 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
```


major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate.";
RT Exp. Parasitol. 67:1-11(1988).
RL -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC
CC EMBL; M37213; AAA29611.1; -.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 136174 MW; 5B59CEFA2F9A026 CRC64;
Query Match
Best Local Similarity 6.8%; Score 103; DB 1; Length 1726;
Matches 57; Conservative 53; Mismatches 99; Indels 92; Gaps 13;
QY 28 NKYLDQVTPWTFBETIKE-LSRFKQEVYSQ-----EASVLVDGIKYL- 68
Db 1322 DEYLDQVVTGSAISVTMDNLSGFENEYDVTYKPLAGVYRSLRKKQIEKNIPTNLFNLND 1381
QY 69 LMDS--QDKYFQATQVYVGVVQVVTQLLSAVILLDFEYNEKKAQAQKILIRLIDGVK 125
Db 1382 ILNSLKKRKTF-----LDVLSDLQMFQKHISNE-----YIIDSFK 1419
QY 126 KLNBAQSKLLTSQSFNASGKLLALDQLTNDF---SEKSSYFQ-----SQ 169
Db 1420 LNSEQKNTLLSKYK-----IKESVENDIKFAQEGISYVEKVLAKYKDDLESIKK 1470
QY 170 VDRIRKEAYAGAAAGVAGPGLAISVSIAGVLEGKLIPEL-----NNELKTQVNF 222
Db 1471 VIKKEKFPSPPTTPSP-----AKTDQKESKFLPPLTNIETLYNNLVNKIDYDL 1524
QY 223 TSLSATVQAKNDIDAALKLKLATIAAIGETKTETTRFYVDYDDLMMLSLKGAQKMI 282
Db 1525 INLRKAKINDCNVKEDEAHVKI-TKLSDLKALDKIDLFQHNDF-----AIKKLI 1574
QY 283 N 283
Db 1575 N 1575
RESULT 26
BPEA HUMAN
ID BPEA HUMAN STANDARD; PRT; 5171 AA.
AC Q94833; Q9N178; Q9N6J3; Q8WKK9; Q96AK9; Q96DQ5; Q9H555;
DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1, isoforms 5/9/10 (Trabeculin-beta)
DE (Bullous pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein)
DE (Dystonia muscularum protein)
GN BPAGI OR DMH OR DT OR KIAA0728.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1] SEQUENCE FROM N.A. (ISOFORM 6), AND TISSUE SPECIFICITY.
RP TISSUE=Keratinocytes;
RC MEDLINE=21839111; PubMed=11751855;
RA Okumura M., Yamakawa H., Ohara O., Owarike K.;
RT "Novel alternative splicings of BPAGI (bullous pemphigoid antigen 1) including the domain structure closely related to MACF (microtubule actin cross-linking factor).";
RL J. Biol. Chem. 277:6682-6687(2002).
RN [2]
RP SEQUENCE OF 1342-5171 FROM N.A. (ISOFORM 10).
RC TISSUE=Ductenium;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whitting M., Madan A., Young A.C., Green E.D., Dickson M.C., Blakesley R.W., Touchman J.W., Green E.D., Myers R.M., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1702-4156 FROM N.A. (ISOFORM 6).
RC TISSUE=Brain, Placenta, and Tongue;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H., Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K., Isogai T., Nishi T., Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T., Irie R., Otsuki T., Sato H., Kikuchi H., Masuho Y., Nagai K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 3535-5171 FROM N.A. (ISOFORM 6).
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [5]
RP SEQUENCE OF 4107-5171 FROM N.A. (ISOFORM 6).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";


```
Db 3401 -----SDITQKVEVDRAILSSQQFDOAADAEALSWITETKTKMLSLGICLELQDQTSACL 3456
QY 265 D-----YDMLSLKGAAGKQKMTNCNYYQQRHGHKTLFEV 300
Db 3457 QVQKTFTEMLHRDIIDLVK-SGHKIMTACSEEEKQMKKKLDKV 3502

RESULT 27
ANT1 ONCVO
ID ANT1 ONCVO STANDARD; PRT; 2022 AA.
AC F21249;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Major antigen.
GN OVTL.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95287898; PubMed=7770081;
RA Triteeraprapab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,
RA Neubert T.A., Scott A.L.;
RT "Molecular cloning of a gene expressed during early embryonic
RT development in Onchocerca volvulus."
RL Mol. Biochem. Parasitol. 69:161-171(1995).
RN [2]
RP SEQUENCE OF 733-866 FROM N.A.
RX MEDLINE=89127417; PubMed=2464764;
RA Donaldson J.E., Duke B.O.L., Moser D., Zeng W., Erundu N.E.,
RA Lucius R., Renz A., Karam M., Flores G.Z.;
RT "Construction of Onchocerca volvulus cDNA libraries and partial
RT characterization of the cDNA for a major antigen."
RL Mol. Biochem. Parasitol. 31:241-250(1988).
CC -!- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U12881; AA80009.1; -.
CC DR EMBL; J03995; AA29412.1; -.
CC DR PIR; T43214; T43214.
CC KW Antigen; Coiled coil.
FT DOMAIN 74 120 COILED COIL (POTENTIAL).
FT DOMAIN 151 251 COILED COIL (POTENTIAL).
FT DOMAIN 327 384 COILED COIL (POTENTIAL).
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;

Query Match 6.8%; Score 102.5; DB 1; Length 2022;
Best Local Similarity 21.1%; Pred. No. 49;
Matches 69; Conservative 50; Mismatches 101; Indels 107; Gaps 13;

QY 7 EQTEVVKSAIETADGALDLYNKYLDQV-----IPWKTDFETIKELSRPKQYSEASVTV 62
Db 649 EQDEIIRKQKLAKELADLENKLNNETFMKGDAEKLQNRHLDEIDNFKQINEYIT--- 705
QY 63 GDIKVLLMDSQDKYFEATQTVVWGVVTVQLLSAVILLFDYNEKKASACKDILIRLDD 122
Db 706 ---EVTIIRQNDPFD-----TQM-----KTNAKLSSMKNLSIAAKKE 741
QY 123 GVKKLNEAQSLLTSSQFNNAAGKLLALDSQNDPSEKSSYFQSQVDRIKRE----- 176
Db 742 -IEKLSEMNRL---QQDKNDLIGAKQKGDTEL-NLLTEKIRKVEIBFERIKKNQLEL 796
```

```
QY 177 -----AYAGAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRL 215
Db 797 HERTARDULKQTNHLLAKLEEARADIVA-----LNDRKL 833
QY 216 -KTVQNFTSLSATVQKANKIDAAK-----LKLATEIAAIGIKTETETTRFV 264
Db 834 AKMDANFKIKLDETIKSPADHETIKSRSEKIKIVKHETIKYINKYRABLE--KLES 891
QY 265 DYDDLMLSLKGAAGKQKMTNCNYYQQR 291
Db 892 DKDDL-----EKRIIGLQDELNEK 910

RESULT 28
Y855 RICPR
ID Y855 RICPR STANDARD; PRT; 520 AA.
AC G9ZG5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP855.
GN RP855.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichterz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ235273; CAA15279.1; -.
CC DR FIR; G71647; G71647.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 337 357 POTENTIAL.
SQ SEQUENCE 520 AA; 61113 MW; A4B4DBA1BE594034 CRC64;

Query Match 6.7%; Score 102; DB 1; Length 520;
Best Local Similarity 20.4%; Pred. No. 9.7;
Matches 42; Conservative 33; Mismatches 77; Indels 54; Gaps 6;

QY 25 DLYNKYLDQVTPWKTDFETIKELSRFKQYSEASVLDGDKVLLMDSQDKYFEATQTVY 84
Db 190 DKYNDYF-QCVYVVSOTSGIKFIKHKN-----NIITRNLDYFPDKTNTV 235
QY 85 EWGVVTVQLLSAVILLFDYNEKKASACKDILIRLDDGVKLNKAQKSLTSSQSFNNA 144
Db 236 Q--GIIEQIHDFFILFKNYISN--LDQRCVILIVDDELKELLE-----STNFEDR 283
QY 145 SGKLLALDSQNDPSEKSSYFQSQVDRIKREAYAGAAAGIVAGPGLIISYSIAAGVIE 204
Db 284 NVIFIPVDNLNKQTLERFIDANISRL-----FLE 315
QY 205 GKLIPELNNRLKTVQNFITSLSATVK 230
Db 316 YKSFAYNNNLSIKKLVTKDITFK 341

RESULT 29
```

```
MYHA_BOVIN
ID MYHA_BOVIN STANDARD; PRT; 1976 AA.
AC Q27991;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
GN MYH10.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohara M., Ishiguro N., Shinagawa M.;
RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE OF 204-302 FROM N.A.
RP TISSUE=Brain cortex;
RC MEDLINE=95301542; PubMed=7782316;
RA Itch K., Adelstein R.S.;
RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle
RT myosin heavy chain II-B";
RL J. Biol Chem. 270:14533-14540(1995).
CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC CAPPING (BY SIMILARITY).
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB022023; BAA36494.1; -
DR EMBL; U15716; AAA87715.1; -
DR HSPSP; P10587; I2R2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PR00612; IQ; 1.
DR Pfam; PR00063; myosin_head; 1.
DR Pfam; PR002736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_Tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50036; IQ; 1.
KW Myosin; ATP-binding; calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 785
FT DOMAIN 786 815
FT DOMAIN 816 855
FT NP_BIND 148 1976
FT ATP (POTENTIAL).
FT MOD_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
FT MOD_RES 1976 AA; 229037 KW; 614435445IC0F790 CRC64;
SQ
Query Match 6.7%; Score 102; DB 1; Length 1976;
Best Local Similarity 21.1%; Pred No. 51;
Matches 56; Conservative 88; Mismatches 88; Indels 68; Gaps 10;
QY 52 QEVQSEASVLVGIKVLMDSDQKYFEATQTVVGVVLTQVLSAVILLDFEYNEKASA 111
DB 972 EKVTAEAKIKKMBEIEILLDDONSKFTKEKLM-----DRIACSSQLAESEKAKNLA 1026
QY 112 ----OKDILIRILDGDKVKNLNEAKSLTSSQSFNNASGKLLALDQLTNDSEKSSYFQ 167
DB 1027 KIRNKQVWISDLERLKKEEKTRQBLEKAKRK-----LDGE-TTDLQDQAEALQ 1075
QY 168 SQVDRI-----RKEAYAGAAAGIVAGPFGLLIISYSIAAGVIEG-----KULPELNNR 214
DB 1076 AQIDELKIQVAKKEEELQGA-----LARGDDETLHKNNALKVYRELQAQ 1119
QY 215 LKTVQNFPTSLSATVKGANKDIDAAKCLKLATEIAAIGETETETETTFYVDYDMLSL 274
DB 1120 IAELOEDFESEKASRNKAEKQ-----KDLSEEEA---LKELEDT-----LD 1160
QY 275 KGAAGKVMINTCNEYQORHGKKTIFE 299
DB 1161 TTAQQELRTREQEVAELKALEE 1185
RESULT 30
MYHA_HUMAN
ID MYHA_HUMAN STANDARD; PRT; 1976 AA.
AC P35580;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
GN MYH10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96023307; PubMed=7499478;
RA Phillips C.L., Yamakawa K., Adelstein R.S.;
RT "Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and
RT analysis of human tissues with isoform-specific antibodies.";
RL J. Muscle Res. Cell Motil. 16:379-389(1995).
RN [2]
RP SEQUENCE OF 63-722 FROM N.A.
RA MEDLINE=91315803; PubMed=1860190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located
RT on different chromosomes.";
RL Circ. Res. 69:530-539(1991).
CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC CAPPING.
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
```

or send an email to license@isb-sib.ch).

```

CC CC EMBL; M69181; AAA99177.1; -
CC DR PIR; A59252; A59252.
CC DR HSSP; P10587; LBR2.
CC DR Genew; HGNC:7568; MYH10.
CC DR GO; GO:0016459; C:myosin; NAS.
CC DR GO; GO:0003779; F:actin binding activity; NAS.
CC DR GO; GO:0005524; F:ATP binding activity; NAS.
CC DR GO; GO:0016288; P:cytokinesis; NAS.
CC DR InterPro; IPR000048; IQ_region.
CC DR InterPro; IPR001609; myosin_head.
CC DR InterPro; IPR004009; Myosin_N.
CC DR InterPro; IPR002928; Myosin_tail.
CC DR InterPro; IPR002017; Spectrin.
CC DR Pfam; PF00612; IQ; 1.
CC DR Pfam; PF00603; myosin_head; 1.
CC DR Pfam; PF02736; Myosin_N; 1.
CC DR Pfam; PF01576; Myosin_tail; 1.
CC DR PRINTS; PR00193; MYOSINHEAD.
CC DR ProDom; PD000355; myosin_head; 1.
CC DR SMART; SM00015; IQ; 1.
CC DR SMART; SM00242; MYSC; 1.
CC DR PROSITE; PS00096; IQ; 1.
CC KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
CC KW Coiled coil; Alkylation; Multigene family.
CC FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
CC FT DOMAIN 786 815 IQ.
CC FT DOMAIN 845 1976 COILED COIL (POTENTIAL).
CC FT NP_BIND 178 185 ATP (POTENTIAL).
CC FT MOD_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).
CC FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
CC SQ SEQUENCE 1976 AA; 228938 MW; 52BB87FF35EA124F CRC64;

Query Match 6.7%; Score 102; DB 1; Length 1976;
Best Local Similarity 21.1%; Pred. No. 51;
Matches 56; Conservative 53; Mismatches 88; Indels 68; Gaps 10;

QY 52 QEYSCEASVLDGKIVLMDSDQKYFEATQVYVCGVVTQLLSAYILLFDEYNEKKASA 111
DB 972 EKVTAEAKIKWEESEILLEQNSKFIKEKIME-----DRIACSSQLAESEKAKNLA 1026
QY 112 ----OKDILIRLDGKVLKNEAQKSLTSSQSFNNASGKLLALDSQITNDFSEKSYFQ 167
DB 1027 KIRNKQEWISLEERLAKBEKTRCELEKAKRK-----LDGE-TTDLQDQIAELQ 1075
QY 168 SOVDRI-----RKAYAGAAGIVAGPFGHIIYSINAGVIEG-----KLIPELNLR 214
DB 1076 AQIDELKLQAKKEBELOQA-----LARGDDETLHKNNALKVVRELQAO 1119
QY 215 LKTQVNFFTSLSATVKQANKDIDAALKLATEIAAIGIKTETETTRFYVDYDDLMLSL 274
DB 1120 IAELOQEDFESEKASRNKAQKQ-----KRDLSBELEA---LKTELEDT-----LD 1160
QY 275 KGAAKKMNNTCNVEYQORHGKTLFE 299
DB 1161 ITAAQOQLRTKREQEVAELKALEE 1185

RESULT 31
P115_MYCGB
ID P115_MYCGB STANDARD; PRT; 982 AA.
AC P47540; Q49301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P115 protein homolog.
GN MG298
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gccayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fritschmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritschmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 915-981 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing";
RL J. Bacteriol. 175:7918-7930(1993).
CC -1- DOMAIN: Consists of two putative central coiled-coil regions
CC flanked by putative globular regions at the N- and C-termini.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SIMILAR TO OTHER MYCOPLASMA
CC P115.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39710; AAC71520.1; -
CC EMBL; U02177; AAD12461.1; -
CC TIGR; MG298; -
CC InterPro; IPR003405; SMC_C.
CC Pfam; PF02483; SMC_C; 1.
CC Pfam; PF02453; SMC_N; 1.
CC Hypothetical protein; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (POTENTIAL).
FT DOMAIN 174 232 COILED COIL (POTENTIAL).
FT DOMAIN 264 383 COILED COIL (POTENTIAL).
FT DOMAIN 573 835 COILED COIL (POTENTIAL).
FT DOMAIN 955 982 COILED COIL (POTENTIAL).
FT DOMAIN 884 912 ALA/ASP-RICH (DA-BOX).
FT CONFLICT 975 981 YVESDS -> ICIWKF (IN REF. 2).
SQ SEQUENCE 982 AA; 111073 MW; E31CA2430B95A87 CRC64;

Query Match 6.7%; Score 101.5; DB 1; Length 982;
Best Local Similarity 21.3%; Pred. No. 23;
Matches 81; Conservative 44; Mismatches 130; Indels 125; Gaps 16;

QY 3 SIFAEQTVVVKSAETADGALDLYNKVLDQVPHKTFDETIKELSRKQESQASVLV 62
DB 149 SMFVEAKPEERRKIFEDASG-IGRYTKREEVY-----NQLNRTLINLQKQ-VSVVLNEUK 201
QY 63 GGIKVLMDSDQ-----KYFEATQVYVCGVVTQLLSAYILLFDEYNEKKASA 111
DB 202 KOLKKTTLQAKAQPFIRVKNELKELEAVLGEVLAQTEL-----DKFNQINSS 253
QY 112 QKDILI-----RILDDGV-----KLNENAKSLTSSQSFNNASGKLLALDSQL 155
DB 254 EHDFKIHEPQELLEBEQVIFNFRSHFADMQSNELQKELQDIYOKINELEQRKVIDQL 313
QY 156 TNDFSEKSSY-----FQSQ-----VDRIRK 175
DB 314 RQFSQKDEKQAAKLLIYVDQTLQDGFENQLNSKTTITDLEKLINEQSLVDQIKL 373
QY 176 EYAGAAGIVAGPFGHIIYSINAGVIEGKLIPEL-NNRLKTVNFFTSLSATVKQANK 234
DB 374 QIEKNTA-----DLIYORSL-----KTIIELOTNELKKTNN-----ANILVKAN- 413

```

QY 235 DIDAAKLKLATIAAIGIKETETTRFYVDYDDMLSL-KGAKKKMTNCE----- 287
 Db 414 -----ALTGIINTLGTFLKDKQYKAILKALGKSGLYLVNNAIAIQAID 460

QY 288 --YQQRHGKTKLFEPDVAS 305
 Db 461 FLVKNIGKVTFLPLDDVAS 480

RESULT 32
 DMD CANFA STANDARD; PRT; 3680 AA.
 AC Q97592;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dystrophin.
 GN DMD.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Golden retriever;
 RA Carville K.S., Mann C.J., Schatzberg S.J., Wilton S.D.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May play a role in anchoring the cytoskeleton to the
 CC plasma membrane.
 CC -!- SUBUNIT: Interacts with the syntrophins SNTA1, SNTB1, SNTB2, SNTG1
 CC and SNTG2 (By similarity).
 CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
 CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
 CC ASP-120, ASP-180, OR BETA-FODRIN).
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -!- SIMILARITY: Contains 22 spectrin repeats.
 CC -!- SIMILARITY: Contains 1 WW domain.
 CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF070485; AAC83646.1; -
 CC HSP; P46939; IQAG.
 DR InterPro; IPR001589; Actbind actnin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001202; WW Rep5_WWP.
 DR InterPro; IPR000433; ZnF_ZZ.
 DR Pfam; PF00307; CH; 2.
 DR Pfam; PF00435; spectrin; 21.
 DR Pfam; PF00397; WW; 1.
 DR Pfam; PF00569; ZZ; 1.
 DR SMART; SMC0033; CH; 2.
 DR SMART; SMC0456; WW; 1.
 DR SMART; SMC0291; ZnF_ZZ; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS00021; CH; 2.
 DR PROSITE; PS01159; WW DOMAIN_1; 1.
 DR PROSITE; PS00020; WW DOMAIN_2; 1.
 DR PROSITE; PS01357; ZF_ZZ_1; 1.
 DR PROSITE; PS01357; ZF_ZZ_2; 1.
 DR PROSITE; PS00135; ZF_ZZ_2; 1.
 KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
 KW Repeat; Zinc-finger.
 FT DOMAIN 15 CH 1.

PT DOMAIN 134 237 CH 2.
 FT REPEAT 340 448 SPECTRIN 1.
 FT REPEAT 449 557 SPECTRIN 2.
 FT REPEAT 560 668 SPECTRIN 3.
 FT REPEAT 720 829 SPECTRIN 4.
 FT REPEAT 831 935 SPECTRIN 5.
 FT REPEAT 944 1047 SPECTRIN 6.
 FT REPEAT 1050 1156 SPECTRIN 7.
 FT REPEAT 1159 1265 SPECTRIN 8.
 FT REPEAT 1268 1363 SPECTRIN 9.
 FT REPEAT 1470 1570 SPECTRIN 10.
 FT REPEAT 1573 1678 SPECTRIN 11.
 FT REPEAT 1681 1782 SPECTRIN 12.
 FT REPEAT 1879 1981 SPECTRIN 13.
 FT REPEAT 2013 2103 SPECTRIN 14.
 FT REPEAT 2106 2210 SPECTRIN 15.
 FT REPEAT 2213 2320 SPECTRIN 16.
 FT REPEAT 2470 2572 SPECTRIN 17.
 FT REPEAT 2575 2681 SPECTRIN 18.
 FT REPEAT 2684 2797 SPECTRIN 19.
 FT REPEAT 2800 2902 SPECTRIN 20.
 FT REPEAT 2904 2926 SPECTRIN 21.
 FT REPEAT 2929 3035 SPECTRIN 22.
 FT DOMAIN WW
 FT ZN_FING 3050 3083
 FT ZN_FING 3302 3349 ZZ-TYPE
 SQ SEQUENCE 3680 AA; 425650 MW; 539FIC9D72377872 CRC64;

Query Match 6.7%; Score 101.5; DB 1; Length 3680;
 Best Local Similarity 18.9%; Pred. No. 1.2e+02;
 Matches 65; Conservative 67; Mismatches 127; Indels 85; Gaps 15;

QY 7 EQTVVVKVSALEADGALDLYNKVLDQVLPWKTFDE---TIKLSRFKQYSO-EASV-- 60
 Db 1159 DKTVSLQKQDSEHHEWNTQAEYLERDFEYKTPDELQTAVEEMKKAQKAEKAKVL 1218
 QY 61 -----LVGDIKVLMLDSQDKYFEATQTVYEW-----CGVVTQLASAYILLFDEYN 105
 Db 1219 LTESVNSVIAQAPPAQAEALKKELDTLTNTYQWLCTRLNGCKTLEBVMACWHELL-SYL 1277
 QY 106 EKASAKQDILIRI-----LDDGVKVLNEAKQLLTSSQSFNNASGKLALDLSQLTND-- 158
 Db 1278 EKANKVSEVEVKLKTENISGGAEEIAEVLDSLEMLMQHSEDPNQIRLAIQTLDDGV 1337
 QY 159 ----FSKSSYFQSDVRIRKEAYAGAAAGIVAGPFLIISYIAAGVIEGKLIPELNRR 214
 Db 1338 MDELNEELETFSRWELHEEAVERRK-----LLEQSIQSA-----QEIEKS 1380
 QY 215 LKTVQNFETLSATVKQ-----ANKDIDAAKLKLATEIAAIGIKETETTRFYVDYDDL 269
 Db 1381 LHLIQE---SLSSIDKQLAAVIAADK-VDAAMQFQEAQ-----KIQSDLTSHSILEEM 1429
 QY 270 -----MLSLKGAKKMINTCNEYQQRHGKTKLFEPV 301
 Db 1430 KKHNGKETAGRVLSQIDVAQKQLQDVSMKFR-----LFQKP 1466

RESULT 33
 HMMR MOUSE STANDARD; PRT; 794 AA.
 ID AC Q00547;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hyaluronan mediated motility receptor (intracellular hyalurononic acid
 DE binding protein) (Receptor for hyaluronan-mediated motility).
 GN HMMR OR IHABP OR RHAMM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
 RC TISSUE=Lung;

RA MEDLINE=98264863; PubMed=9601097;
RA Hofmann M., Fieber C., Assmann V., Goettlicher M., Sleeman J.,
RA Plug R., Howells N., von Stein O., Ponta H., Herrlich P.;
RT "Identification of IHABP, a 95 kDa intracellular hyaluronate binding
RT protein";
RL J. Cell Sci. 111:1673-1684 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Zhao Y., Zhang S., Turley E.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-183 FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=129/SV;
RX MEDLINE=98107769; PubMed=9889313;
RA Fieber C., Plug R., Sleeman J., Dall P., Ponta H., Hofmann M.;
RT "Characterization of the murine gene encoding the intracellular
RT hyaluronan receptor IHABP";
RL Gene 226:41-50 (1999).
RN [4]
RP SEQUENCE OF 164-794 FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=BALB/c; TISSUE=fibroblast;
RX MEDLINE=96011639; PubMed=7590272;
RA Entwistle J., Zhang S., Yang B., Wong C., Li Q., Hall C.L., A J.,
RA Mowat M., Greenberg A.H., Turley E.A.;
RT "Characterization of the murine gene encoding the hyaluronan receptor
RT RHAMM";
RL Gene 163:233-238 (1995).
RN [5]
RP SEQUENCE OF 318-794 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=BALB/c;
RX MEDLINE=92299690; PubMed=1376732;
RA Hardwick C., Hoare K., Owens R., Hohn H.P., Hook M., Moore D.,
RA Cripps V., Austen L., Nance D.M., Turley E.A.;
RT "Molecular cloning of a novel hyaluronan receptor that mediates tumor
RT cell motility";
RL J. Cell Biol. 117:1343-1350 (1992).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=94308286; PubMed=7518470;
RA Hall C.L., Wang C., Lange L.A., Turley E.A.;
RT "Hyaluronan and the hyaluronan receptor RHAMM promote focal adhesion
RT turnover and transient tyrosine kinase activity";
RL J. Cell Biol. 126:575-588 (1994).
RN [7]
RP ERK REGULATION, AND SUBCELLULAR LOCATION.
RX MEDLINE=98225222; PubMed=9556628;
RA Zhang S., Chang M.C., Zylka D., Turley S., Harrison R., Turley E.A.;
RT "The hyaluronan receptor RHAMM regulates extracellular-regulated
RT kinase";
RL J. Biol. Chem. 273:11342-11348 (1998).
RN [8]
RP REVIEW.
RX MEDLINE=99059494; PubMed=9845361;
RA Hofmann M., Assmann V., Fieber C., Sleeman J.P., Moll J., Ponta H.,
RA Hart I.R., Herrlich P.;
RT "Problems with RHAMM: a new link between surface adhesion and
RT oncogenesis?";
RL Cell 95:591-592 (1998).
RN [9]
RP FUNCTION: INVOLVED IN CELL MOTILITY. WHEN HYALURONAN BINDS TO
CC HMGR, THE PHOSPHORYLATION OF A NUMBER OF PROTEINS, INCLUDING THE
CC FOCAL ADHESION KINASE OCCURS. MAY ALSO BE INVOLVED IN CELLULAR
CC TRANSFORMATION AND METASTASIS FORMATION, AND IN REGULATING
CC EXTRACELLULAR-REGULATED KINASE (ERK) ACTIVITY.
CC -1- SUBUNIT: SUBUNIT OF THE HARC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE AND CYTOPLASMIC.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=RHAMM1V4;
CC IsoId=Q00547-1; Sequence=Displayed;
CC Name=RHAMM1;
CC IsoId=Q00547-2; Sequence=VSP 004287;
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed.
CC [9]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC [10]
CC EMBL: AF031932; AAC12655.1; -;
DR EMBL: AF079222; AAD0670.1; -;
DR EMBL: AJ005919; CAA06768.1; -;
DR EMBL: AJ005920; CAA06768.1; JOINED.
DR EMBL: AJ005921; CAA06768.1; JOINED.
DR EMBL: AJ005922; CAA06768.1; JOINED.
DR EMBL: AJ005923; CAA06768.1; JOINED.
DR EMBL: AJ005924; CAA06768.1; JOINED.
DR EMBL: X64550; CAA45849.1; -;
DR EMBL: X64550; CAA45848.1; -;
DR PIR: JC4298; JC4298.
DR MGD: MGI:104667; Hmnr.
DR Hyaluronic acid; Alternative splicing; Repeat; Glycoprotein.
FT DOMAIN 442 546
FT REPEAT 442 462
FT REPEAT 463 483
FT REPEAT 484 504
FT REPEAT 505 525
FT REPEAT 526 546
FT DOMAIN 719 729
FT DOMAIN 741 750
FT CARBOHYD 53 53
FT CARBOHYD 134 134
FT CARBOHYD 279 279
FT CARBOHYD 446 446
FT CARBOHYD 467 467
FT CARBOHYD 488 488
FT CARBOHYD 509 509
FT CARBOHYD 530 530
FT CARBOHYD 561 561
FT CARBOHYD 601 601
FT VARSPLIC 218 242
FT CONFLICT 19 19
FT CONFLICT 55 55
FT CONFLICT 71 71
FT CONFLICT 89 91
FT CONFLICT 94 94
FT CONFLICT 540 540
FT CONFLICT 668 668
SQ SEQUENCE 794 AA; 91799 MW; 74DB3D236224499C CRC64;
Query Match 6.7%; Score 101; DB 1; Length 794;
Best Local Similarity 20.1%; Pred. No. 19;
Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;
QY 1 MTSFAEQVVEVVKGAETADGALDLYNKYLDQVTPWKTFDTIKELSRFKQEVQSEASV 60
DB 416 LNNLLREKEVELEKHAHAQAALLIAQEKYNDTA---QSLRDVTAQESVQEKYNDTAQS 472
QY 61 LVGDIKVLMSQDKYFEATQVYEWCGVVTQLLSAYILLFVEYNEKASAKDILIRL 120
DB 473 L-RDVTAQLESEQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQLSRDVTVAQ-L 522
QY 121 DGVGKLNBAQKSLTSS---QSFNNASGKLIALDSQLTN-DFSEKSSYFQSQVDRIKE 176
DB 523 ESQVEKYNDTAQSLRDVSAQLESYKSTLKEIE-DLKENLTLOEKVMAEKSVEDVQQQ 581
QY 177 AVAGAAAGIVAGPFGILIIISYTAAGVIEGKLIPELNLR-----LKTQVNF-----TSL 225
DB 582 ILTAESTN-----QEYA-----RMVQDLQNRSLKEEIEKITSSFLKIDTL 624
QY 226 SATVQKANKDI-----DAAKLKLATEIA-----AIGIKTETETTRVVD-Y 266
DB 625 KNOLRQDQDFRKLQLEKGRKTAENVTMTMINKWRLLYELYEKTFPQQQLDAF 684

KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding; Polymorphism.
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IQ.
FT DOMAIN 842 1938 COILED COIL (POTENTIAL).
FT NP BIND 178 185 ATP.
FT DOMAIN 557 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT DOMAIN 790 807 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 816 833 CALMODULIN-BINDING (BY SIMILARITY).
FT MOD RES 129 129 METHYLATION (TRI-) (POTENTIAL).
FT MOD RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
FT MOD RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
FT VARIANT 194 194 Y -> D.
FT VARIANT 545 545 S -> A.
FT VARIANT 838 838 I -> S.
SQ SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;
Query Match 6.7%; Score 101; DB 1; Length 1938;
Best Local Similarity 22.5%; Pred. No. 58;
Matches 75; Conservative 44; Mismatches 128; Indels 86; Gaps 13;
QY 5 FAEQTVVVKAGIETADGALDLYNKYLDQVTPWTFETIKELSRFQYEQSASVLVGD 64
Db 1173 FQMRERDEATLQHEATAALRKHADSV---AELGEQIDNLRQVKLEKES-----E 1225
QY 65 IKVLLMDSQDYFEATQVWCGVVTQLLSAYILLFDEYNEKASAKQDILILQDGV 124
Db 1226 FLELDD-----VTSNMQII-----KAKANLEKVSITLEDQA 1258
QY 125 K-----KLANEAKSL---LTSSQSFNNAGSKLL-----ALDSQJTNDFEKSYSFQSV 170
Db 1259 NEYRVKLEAQRSLNDFDTQAKLQTEGELARQLEKEALISQITRG---KLSYTOQME 1315
QY 171 DRIKEAYAGAAAGVAPFGILLISYSAAGVIEGKLIPE-----LNRL 215
Db 1316 DKRLQLEEGKAKNALA-----HALQSRHCDLLREYEEEMAKELQRLVSKAN 1367
QY 216 KTVQNFTLSATVQKANKDIDAKLKLATE-----IAAIGIKTETET--RFVVDY 265
Db 1368 SEVAQWRKYETDAIQTTELEEAQKQKLAQRLQDAEAVEAVNAKSSLEKTKHLQNEI 1427
QY 267 DDLMLSLKGRAKKMINTCNEYQORHGKTLIFE 299
Db 1428 EDLMVDVRSNAAA---AALDKKORNFDKILAE 1457
RESULT 36
MYH6 HUMAN
ID MYH6 HUMAN STANDARD; PRT; 1939 AA.
AC P13533; Q13943; Q14906; Q14907;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
GN MYH6 OR MYHCA
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92133665; PubMed=1776652;
RX Matsuda R., Beisel K.W., Furutani M., Arai S., Takao A.;
RT "Complete sequence of human cardiac alpha-myosin heavy chain gene and
RT amino acid comparison to other myosins based on structural and
RT functional differences";
RL Am. J. Med. Genet. 41:537-547(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94140346; PubMed=8307559;
RA Epp T.A., Dixon I.M., Wang H.Y., Sole M.J., Liew C.-C.;

RT "Structural organization of the human cardiac alpha-myosin heavy
RT chain gene (MYH6).";
RL Genomics 18:505-509(1993).
RN [3]
RP SEQUENCE OF 1-177 AND 1551-1939 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN [5]
RP SEQUENCE OF 1407-1939 FROM N.A.
RX MEDLINE=89299163; PubMed=2969919;
RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
RT "Molecular cloning and characterization of human cardiac alpha- and
RT beta-form myosin heavy chain complementary DNA clones. Regulation of
RT expression during development and pressure overload in human
RT atrium";
RL J. Clin. Invest. 82:524-531(1988).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00943; BAA00791.1; -
CC EMBL; Z20656; CAA79675.1; -
CC EMBL; M25140; AAA60386.1; -
CC EMBL; M25162; AAA60386.1; JOINED.
CC EMBL; M25142; AAA60387.1; -
CC EMBL; M25141; AAA60387.1; JOINED.
CC EMBL; M21664; AAA36344.1; -
CC PIR; A46762; A46762.
CC HSP; P08739; LMND.
CC Genew; HGNC:7576; MYH6.
CC MIM; 160710; -
CC GO; GO:0005859; Cmuscle myosin; TAS.
CC GO; GO:0003779; F-actin binding activity; NAS.
CC GO; GO:0005524; F-ATP binding activity; NAS.
CC GO; GO:0005516; F-actin binding activity; NAS.
CC GO; GO:0003776; F-muscle motor activity; NAS.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR001609; myosin_head.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF02126; Myosin_N; 1.
CC Pfam; PF01576; Myosin_tail; 1.


```

FT VARIANT 311 311 I -> V.
FT FTID=VAR_003858.
FT VARIANT 329 329 I -> T.
FT FTID=VAR_003859.
FT VARIANT 357 357 N -> Y.
FT FTID=VAR_003860.
FT VARIANT 489 489 E -> Q (IN EHK).
FT FTID=VAR_003861.
FT VARIANT 536 536 G -> C.
FT FTID=VAR_003862.
FT VARIANT 559 555 MISSING (IN ALIELE 1B).
FT FTID=VAR_003864.
FT VARIANT 632 632 R -> K (IN dSNP:14024).
FT FTID=VAR_003863.
SQ SEQUENCE 643 AA; 65886 MW; DF945DC462257850 CRC64;

Query Match 6.6%; Score 100.5; DB 1; Length 643;
Best Local Similarity 22.2%; Pred.No.16;
Matches 72; Conservative 58; Mismatches 124; Indels 71; Gaps 15;

QY 5 FAEQVEVYVKS---AETADGALDLYNKYLDQVDPKTFDETIKELSRFQKQYQASVL 61
DQ 199 FLEQQNQVLTQWELLQVDTSTRHN-----LEPY--FESFINLRERVQLKSDQSRL 251
QY 62 VGDIK---VLLMDSQKYPE---ATQVYVCGVVTOLLSAYILLFDEYNEKKSAAQKD 114
DQ 252 DSELKMQMDVRYNRYKDEINKRTNAENEFVTKVDVGAYMTKVD-LQAKLDNLQOE 310
QY 115 ILRLDDGVKLNKAQKSLTSSQSFNNA-----SGKLLALDSQL-----TNDFSKS 163
DQ 311 I-----DELTALYQAELSQMTQISETNVILSMNDRSLDLSIIAEVKAQNEIDIAQS 364
QY 164 -----SYFQSOVDRIKEAYAGAAQIVAGPFG-----LIISYIAAGVIEGKLPELNN 213
DQ 365 KAEASLYQSKYEELQ-----ITAGRHGDSVNRKSIIEINRVIQ-RURSEIDN 413
QY 214 RLKTVQNFYTSATYKQAKDIDAKLKLATEIAAIGETETETTFYDYDDLM--- 270
DQ 414 VKQISNLQOS-SDAQRGENALKDAKKNLNDLEDALQAK--EDLARLLRDYQELMNTK 471
QY 271 -----LSLLKGAANKWINTC 285
DQ 472 LALDLLEIATYRTLLGESESRMSGEC 496

RESULT 39
TEAL SCHPO STANDARD; PRT; 1147 AA.
AC P87061;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tip elongation aberrant protein 1 (Cell polarity protein teal).
GN TEAL OR SPCC1223.06.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97344085; PubMed=9200612;
RA Wata J., Nurse P.;
RT teal and the microtubular cytoskeleton are important for generating
RT global spatial order within the fission yeast cell.;
RL Cell 89:939-949(1997).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

```

```

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Geble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Keather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Woltgens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -|- FUNCTION: CELL POLARITY PROTEIN. MAY ACT AS AN END MARKER.
CC DIRECTING THE GROWTH MACHINERY TO THE CELL POLES. MAY ALSO
CC INFLUENCE MICROTUBULAR ORGANIZATION, AFFECTING THE MAINTENANCE OF
CC A SINGLE CENTRAL AXIS.
CC -|- SUBCELLULAR LOCATION: PRESENT AT BOTH POLES OF THE CELL THROUGHOUT
CC THE CELL CYCLE WHETHER THEY ARE GROWING OR NOT. LOCATED AT THE
CC ENDS OF MICROTUBULES GROWING TOWARDS THE CELL POLES. AN INTACT
CC MICROTUBULAR SKELETON IS REQUIRED TO MAINTAIN THE LOCATION AT THE
CC CELL TIPS; ON THE OTHER HAND, A NORMALLY ORGANIZED ACTIN
CC CYTOSKELETON IS NOT REQUIRED.
CC -|- SIMILARITY: Contains 5 Kelch repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y12709; CAA73245.1; -.
CC EMBL; AL031579; CAA20875.1; -.
CC PIR; T40866; T40866.
CC GeneDB SPombe: SPCC1223.06; -.
CC InterPro: IPR006652; Kelch_rep.
CC Pfam; PF01344; Kelch; 5.
CC SMART; SM00612; Kelch; 3.
CC REPEAT 94 144 KELCH 1.
CC REPEAT 146 198 KELCH 2.
CC REPEAT 254 303 KELCH 3.
CC REPEAT 305 351 KELCH 4.
CC REPEAT 355 402 KELCH 5.
CC REPEAT 611 649 COILED COIL (POTENTIAL).
CC DOMAIN 716 838 COILED COIL (POTENTIAL).
CC DOMAIN 879 990 COILED COIL (POTENTIAL).
CC DOMAIN 1084 1105 COILED COIL (POTENTIAL).
CC SEQUENCE 1147 AA; 127436 MW; 7BE65F6C666EF4F8 CRC64;

Query Match 6.6%; Score 100.5; DB 1; Length 1147;
Best Local Similarity 23.1%; Pred.No.33;
Matches 76; Conservative 44; Mismatches 118; Indels 91; Gaps 16;

QY 7 EQVEVYKSAIETADGALDLYNKYLDQVDPKTFDETIKELSRFQKQYQASVL 56
DQ 761 DQVTVINKFAFER-----DQFRSRM---CFENTIKDLTR-KMEATDMLNVLHESL 808
QY 57 -----EASVLGVGIKVLMD--SQDKYFATQVYVCGVVTOLLSAYILLFDEYNEKKA 109
DQ 809 RSVOTENSELVTETALLKAEVLKVKQAIIDANANIYD-----KLTADHTNYETV 856

```

QY 110 SAQKDILIRLDDGVKXKNAQKSLTSSQSFNNASGKLALDSQLTNDFSEK-----162
 Db 857 SA-----DINQNKETLDKLNGSSDFKKNTEILLHQDIRITNAKLEKREKLINA 906
 QY 163 SSVFQSQVDRIRKAYAGA--AGIVAGPGLIISVIAAGVIEGKLIPLNRLKTVQN 220
 Db 907 SKYIE---DTRSEIQBAEKVSNLPSNFKENSMQQLM-KALEQRNTGAKOLVN 962
 QY 221 FFTSLSATVQKANDIDAALKLATEIAAIGEIKTETETTRFFVYVDLMLSLKGAAXK 280
 Db 963 LRMQLST---ATSELDMLKLKRTTALAEESP-----DYSDI-LSILRADMS 1008
 QY 281 -----MINTCNFYQHQHKKTLFE 299
 Db 1009 FHLHKGQGVLLDTLNGVKGFG---IFE 1034

RESULT 40

SMC2_MOUSE STANDARD; PRT: 1191 AA.
 AC Q8CG48; Q61076; Q9CS17; Q9CSD8;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Structural maintenance of chromosome 2-like 1 protein (Chromosome-
 associated protein B) (XCAP-B homolog) (FGF-inducible protein 16).
 GN SMC2L1 OR SMC2 OR CAPE OR FIN16.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cobbe N., Heck M.M.S.;
 RL "Phylogenetic analysis of SMC proteins";
 RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE OF 1-284 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauber E., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boifelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Haegawa Y., Kawai H., Kottseki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690 (2001).
 RN [3]
 RP SEQUENCE OF 801-1191 FROM N.A.
 RX MEDLINE=96226358; PubMed=9649829;
 RA Guthridge M.A., Seldin M., Basilico C.;
 RT "Induction of expression of growth-related genes by FGF-4 in mouse
 fibroblasts";
 RL Oncogene 12:1267-1278 (1996).

CC -!- FUNCTION: Central component of the condensin complex, a complex
 CC required for conversion of interphase chromatin into mitotic-like
 CC condense chromosomes. The condensin complex probably introduces
 CC positive supercoils into relaxed DNA in the presence of type I
 CC topoisomerases and converts nicked DNA into positive knotted forms

CC in the presence of type II topoisomerases (By similarity).
 CC -!- SUBUNIT: Forms an heterodimer with SMC4L1. Component of the
 CC condensin complex, which contains the SMC2L1 and SMC4L1
 CC heterodimer, and three non SMC subunits that probably regulate the
 CC complex: BRRN1/CAPH, CNAPI/CAPD2 and CAPG (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. In interphase
 CC cells, the majority of the condensin complex is found in the
 CC cytoplasm, while a minority of the complex is associated with
 CC chromatin. A subpopulation of the complex however remains
 CC associated with chromosome foci in interphase cells. During
 CC mitosis, most of the condensin complex is associated with the
 CC chromatin. At the onset of prophase, the regulatory subunits of
 CC the complex are phosphorylated by Cdc2, leading to condensin's
 CC association with chromosome arms and to chromosome condensation.
 CC Dissociation from chromosomes is observed in late telophase (By
 CC similarity).
 CC -!- DOMAIN: The hinge domain, which separates the large intramolecular
 CC coiled coil regions, allows the heterodimerization with SMC4L1,
 CC forming a V-shaped heterodimer (By similarity).
 CC -!- SIMILARITY: Belongs to the SMC family. SMC2 subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AJ534939; CAD59182.1; -;
 EMBL; AK013109; BAB28654.1; -;
 EMBL; AK019977; BAB31946.1; -;
 EMBL; U42385; AAB08867.1; ALT_INIT.
 DR MGI:106067; SMC2L1.
 DR InterPro: IPR003405; SMC_C.
 DR InterPro: IPR003395; SMC_N.
 DR Pfam: PF02483; SMC_C; 1.
 DR Pfam: PF02463; SMC_N; 1.
 KW DNA condensation; Mitosis; Cell cycle; ATP-binding; Coiled coil;
 KW Nuclear protein.
 FT NP BIND 32 39 ATP (POTENTIAL).
 FT DOMAIN 173 507 COILED COIL (POTENTIAL).
 FT DOMAIN 508 671 FLEXIBLE HINGE.
 FT DOMAIN 672 936 COILED COIL (POTENTIAL).
 FT DOMAIN 963 1031 COILED COIL (POTENTIAL).
 FT DOMAIN 1085 1120 ALA/ASP-RICH (DA-BOX).
 FT CONFLICT 62 62 F -> L (IN REF. 2).
 SQ SEQUENCE 1191 AA; 134272 MW; 582C937D019FD893 CRC64;

Query Match 6.6%; Score 100.5; DB 1; Length 1191;
 Best Local Similarity 20.9%; Pred. No. 34;
 Matches 71; Conservative 48; Mismatches 126; Indels 95; Gaps 12;

QY 1 MTSIFAQTVVYVYKGAISTADGALDLYNKYLDQVTPWKTFTDTIKELSRFKQVYQSEASV 60
 Db 150 IYKVLNMPPEILSMTEBAAGTRVVEYKIAAQ---KTIKKKAKLEIKTILEEITP 205
 QY 61 LVGDIKVLMSDQDKYFQATQVTVWGVVQLLSAYI-----LUFDEYNEKKASAKQDI- 115
 Db 206 TIQIKL----EPRSSYLEYQKVMRE----IEHLSRLYAYQFLRAEDTKERSAGELKEMQ 257
 QY 116 ----LIRLDDGVKKLN-----EAQKSLTSSQSF 141
 Db 258 DQVNLQEVLSENEKKIKALNCEIEELERRKDKETGGKGLKLEDAQAR-VNTKSUSA 316
 QY 142 NNASGKLIALD---SQTNDPSEKSYVFSQVDRIRK-----EAYAGA 181
 Db 317 FDLKKKNLASEETKRLQNSMAEDSKALAAKEVKKITDGLHGLQEASNKDAEALAA 376
 QY 182 AGIVAGPGLIISVIAAGVIEGKLIPLN-----RLKTVQNFITSLSAT 228
 Db 377 QOHFNVAAGUSSNEDGAEATLAGMIACKNDISKAQTEAKQAQMKLKHAAQELKSKQAE 436

```
QY 229 VKQAN-----KQIDA-----AKLKLATEIAAIGIKTETE 258
Db 437 VKQNDGYKKDQDAFEAVKAKKLEKLETKMLNKEENKEE 476

RESULT 41
SBCC LACUA
ID SBCC LACUA STANDARD; PRT; 1046 AA.
AC Q9CF20;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nuclease sbccD subunit C.
GN SBCC OR LL1321.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LL1403.
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753 (2001).
CC -!- FUNCTION: sbccD cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (By similarity).
CC -!- SUBUNIT: Heterodimer of sbcc and sbcd (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC -----
CC EMBL; AB006364; AK05419.1; -.
CC F1R; AB8790; AB8790.
CC DR InterPro; IPR003439; ABC transporter.
CC DR Hydroxylase; Nuclease; Exonuclease; Endonuclease; DNA replication.
CC KW DNA recombination; ATP-binding; Coiled coil; Complete proteome.
CC FT NP BIND 34 41 ATP (POTENTIAL).
CC FT DOMAIN 223 239 COILED COIL (POTENTIAL).
CC FT DOMAIN 268 432 COILED COIL (POTENTIAL).
CC FT DOMAIN 468 505 COILED COIL (POTENTIAL).
CC FT DOMAIN 543 867 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 1046 AA; 120199 MW; 8F7D0D0AC28F8691 CRC64;

Query Match 6.6%; Score 100; DB 1; Length 1046;
Best Local Similarity 22.3%; Pred. No. 32;
Matches 69; Conservative 36; Mismatches 114; Indels 90; Gaps 12;

QY 7 EQTEVW-----KGAETAD---GALDLYNKYLQVLPWK-PDETTELSPRQEVSQ 56
Db 260 EQPEKVISIEIAYKSALENOFNLNNLNDYQOINIEKSAIFENSEYKLLKWAFFL 319
QY 57 EASV--LVGDIKVLMDSDQKFEATQTVYVCGVTVLLSAYILL-----F 101
Db 320 KESIHELDQDK-----QSKSVQNNIAGIISSEKAKDEILLKLTKEKDLNQK 368
QY 102 DEYNEKKAQAQDILIRI-----LDGVKKLNEAQSLLTSSQSNFASGKLALDS 153
Db 369 ENINENKVAEK-IFQIQLSLQVKKQKVELKLEQADNLTLLESFKANLGOAENIS 427
QY 154 QLNDPSEKSVFQSOVDRIKAEYAGAAAGVAGFGLIISYSAAGVIEKGLPELN 213
Db 428 TLQDDVISDDYFINKREERQLEL-----TFRGKLIP----- 459

214 RLKTVQNFPTSLSATVKQANKDIDAQKLAETIAAIGIKTETETTTTPYVDYDDMLSL 273
460 ---TFQK-----VHSHKDDIVGLKELKENTALSSENKDNLEKAKFAYN-----EK 502

QY 274 LKGAAKKMI 282
Db 503 LKGRRLIMI 511

RESULT 42
MYH9 HUMAN
ID MYH9 HUMAN STANDARD; PRT; 1960 AA.
AC P35579; O60805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591308;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.B., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Macann O.T.,
RA McElvyn J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.B., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kerton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissee S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wansley A., Wohldmann P., Pepin K., Neilson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P., Payard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wikinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22."
RL Nature 402:489-495 (1999).
RN [2]
RP SEQUENCE OF 1-1337 FROM N.A.
RX MEDLINE=92003935; PubMed=1912569;
```


Db 1166 EOEVLNKKTBEEAKTHEAQIQEMRQKHSQAV--EELAEQLEQTKRVKANLEKAKQTL 1223
QY 56 QEASVLVGDITKVLMDSDQKQYFEATQTVYEWGVTQQLLSAYILLFDFEYNEKKAQAORDI 115
Db 1224 NERGELANEVKKVLQKGD-----SEKRRKKVBAQLOE 1256
QY 116 LIFILDGVKKLNEAKSLLTSQSFFNNSAGKLLAD---SOLTNDFFSEKSYFQSQVDR 172
Db 1257 LQVKNFEGEVRTELADKVTKLQVELDNTGLSQSDSKSKLTKDFSALQSOLQDTQEL 1316
QY 173 IRKEAVAGAAAGIVAGPFGLLISYSTAAGVIEGKLTPELNNRLKTVQNFETLSATVKA 232
Db 1317 LQEN-----RQKL--SLSTKLQVDEKNSFRQLEEE 1349
QY 233 NKDIDAUKLATEIAAIGETETTFYVDYDL-MLSLKGAQKQVNTCNEYQOR 291
Db 1349 EBAKHNLEKQIATLHAQVADMKKME-----DSVGCLETAEEVKRLQKLDLEGLQR 1400
QY 292 HGKK 295
Db 1401 HEK 1404

RESULT 43
ID DESP HUMAN STANDARD; PRT; 2871 AA.
AC P15924; Q75933; Q14189; Q9UHN4;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Desmoplakin (DP) (250/210 kDa paraneoplastic pemphigus antigen).
GN DSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM DPI).
RC TISSUE=Foreskin;
RX MEDLINE=92115697; PubMed=1731325;
RA Virata M.L.A., Wagner R.M., Parry D.A.D., Green K.J.;
RT "Molecular structure of the human desmoplakin I and II amino
terminus";
RL Proc. Natl. Acad. Sci. U.S.A. 89:544-548(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM DPI).
RA Phillips S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1120-2871 FROM N.A. (ISOFORM DPI).
RC TISSUE=Foreskin;
RX MEDLINE=90153880; PubMed=1689290;
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
RA Angst B.D., Nilles L.A.;
RP "Structure of the human desmoplakins. Implications for function in
the desmosomal plaque";
RL J. Biol. Chem. 265:2603-2612(1990).
RN [4]
RP ARATUM.
RX MEDLINE=90361712; PubMed=2391353;
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
RA Angst B.D., Nilles L.A.;
RL J. Biol. Chem. 265:11406-11407(1990).
RN [5]
RP SEQUENCE OF 2854-2871 FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=20062965; PubMed=10594734;
RA Whitlock N.V., Ashton G.H., Dopping-Hepenstal P.J., Gratian M.J.,
RA Keane F.M., Eady R.A.J., McGrath J.A.;
RT "Striate palmoplantar keratoderma resulting from desmoplakin
haploinsufficiency";
RL J. Invest. Dermatol. 113:940-946(1999).
RN [6]

RP CHARACTERIZATION.
RX MEDLINE=98012209; PubMed=9348293;
RA Kowalczyk A.P., Bornslaeger B.A., Borgwardt J.E., Palka H.L.,
RA Dhaliwal A.S., Corcoran C.M., Denning M.F., Green K.J.;
RT "The amino-terminal domain of desmoplakin binds to plakoglobin and
clusters desmosomal cadherin-plakoglobin complexes.";
RL J. Cell Biol. 139:773-784(1997).
RN [7]
RP VARIANT ARVD8 ARG-299.
RX MEDLINE=22285852; PubMed=12373648;
RA Rampazzo A., Nava A., Malacrida S., Befagna G., Baucce B., Rossi V.,
RA Zimbello R., Simonati B., Basso C., Thiene G., Towbin J.A.,
RA Danieli G.A.;
RT "Mutation in human desmoplakin domain binding to plakoglobin causes a
RT dominant form of arrhythmic right ventricular cardiomyopathy.";
RL Am. J. Hum. Genet. 71:1200-1206(2002).
RN [8]
RP VARIANTS SFVHS LYS-287 AND CYS-2366.
RX MEDLINE=21830938; PubMed=11841538;
RA Whitlock N.V., Wan H., Morley S.M., Garzon M.C., Kristal L., Hyde P.,
RA McLean W.H.I., Pulkkinen L., Uitto J., Christiano A.M., Eady R.A.J.,
RA McGrath J.A.;
RT "Compound heterozygosity for non-sense and mis-sense mutations in
RT desmoplakin underlies skin fragility/woolly hair syndrome.";
RL J. Invest. Dermatol. 118:232-238(2002).
CC -!- FUNCTION: Major high molecular weight protein of desmosomes.
CC Involved in the organization of the desmosomal cadherin-
CC plakoglobin complexes into discrete plasma membrane domains and in
CC the anchoring of intermediate filaments to the desmosomes.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Innermost portion of the desmosomal plaque.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=DPI; Synonyms=DPI;
CC IsoId=P15524-1; Sequence=Displayed;
CC Name=DPII; Synonyms=DF2;
CC IsoId=PF1524-2; Sequence=VSP_005070;
CC -!- TISSUE SPECIFICITY: Isoform VSP is apparently an obligate
CC constituent of all desmosomes; Isoform DPII resides predominantly
CC in tissues and cells of stratified origin.
CC -!- DOMAIN: The N-terminal region is required for localization to the
CC desmosomal plaque and interacts with the N-terminal region of
CC plakophilin 1. The C-terminal region interacts with intermediate
CC filaments.
CC -!- DISEASE: Defects in DSP are the cause of familial arrhythmogenic
CC right ventricular dysplasia-8 (ARVD8) [MIM:607450], an autosomal
CC dominant disorder.
CC -!- DISEASE: Defects in DSP are a cause of striate palmoplantar
CC keratoderma II (PKPS2, KPS2 or SPPK2), characterized by skin
CC thickening in the palms (linear pattern) and the soles (island-
CC like pattern) and flexor aspect of the fingers; and rarely by
CC abnormalities of the nails, the teeth and the hair.
CC -!- DISEASE: Defects in DSP are the cause of skin fragility-woolly
CC hair syndrome (SFVHS) [MIM:607655], an autosomal recessive
CC genodermatosis characterized by focal and diffuse palmoplantar
CC keratoderma, hyperkeratotic plaques on the trunk and limbs, and
CC woolly hair with varying degrees of alopecia.
CC -!- SIMILARITY: Contains 17 plectin repeats.
CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M77830; AAA85135.1; -;
DR EMBL; AL031058; CAAL19927.1; -;
DR EMBL; J05211; AAA35766.1; -;
DR EMBL; AF139065; AAF19785.1; -;

DR PIR; A38194; A38194.
DR PDB; 1LM5; 31-JUL-02.
DR PDB; 1LM7; 31-JUL-02.
DR Genew; HGNC:3052; DSP.
DR MIM; 125647; -.
DR MIM; 607450; -.
DR MIM; 607655; -.
DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
DR GO; GO:0008544; P:epidermal differentiation; TAS.
DR InterPro; IPR001101; Plectin repeat.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00681; Plectin; 10.
DR SMART; SM00250; PLEC; 17.
DR Repeat; Coiled coil; Phosphorylation; Cytoskeleton;
KW Structural protein; Alternative splicing; Disease mutation;
KW 3D-structure.
FT DOMAIN 1 1056 GLOBULAR 1.
FT DOMAIN 1057 1945 CENTRAL FIBROUS ROD DOMAIN.
FT DOMAIN 1946 2871 GLOBULAR 2.
FT REPEAT 347 447 SPECTRIN 1.
FT REPEAT 858 952 SPECTRIN 2.
FT DOMAIN 1018 1945 COILED COIL (POTENTIAL).
FT REPEAT 2009 2045 PLECTIN 1.
FT REPEAT 2046 2083 PLECTIN 2.
FT REPEAT 2084 2121 PLECTIN 3.
FT REPEAT 2122 2159 PLECTIN 4.
FT REPEAT 2163 2197 PLECTIN 5.
FT REPEAT 2198 2233 PLECTIN 6.
FT REPEAT 2231 2288 PLECTIN 7.
FT REPEAT 2289 2326 PLECTIN 8.
FT REPEAT 2327 2364 PLECTIN 9.
FT REPEAT 2365 2402 PLECTIN 10.
FT REPEAT 2406 2440 PLECTIN 11.
FT REPEAT 2456 2493 PLECTIN 12.
FT REPEAT 2507 2544 PLECTIN 13.
FT REPEAT 2610 2647 PLECTIN 14.
FT REPEAT 2648 2685 PLECTIN 15.
FT REPEAT 2724 2761 PLECTIN 16.
FT REPEAT 2762 2799 PLECTIN 17.
FT DOMAIN 2824 2847 6 X 4 AA TANDEM REPEATS OF G-S-R-[SR].
FT DOMAIN 1 584 INTERACTS WITH PLAKOPHILIN 1 AND JUNCTION
FT MOD_RES 2849 2849 PLAKOGLOBIN.
FT VARSP_LIC 1195 1794 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT VARIANT 287 287 Missing (in isoform DP11).
FT VARIANT 287 287 /FTID=VSP 005070.
FT VARIANT 287 287 N -> K (in SFVHS).
FT VARIANT 287 287 /FTID=VAR 015559.
FT VARIANT 287 287 S -> R (in ARVDB).
FT VARIANT 287 287 /FTID=VAR 015402.
FT VARIANT 287 287 R -> C (in SFVHS).
FT CONFLICT 905 905 R -> A (in REF. 2).
FT CONFLICT 1120 1120 D -> R (in REF. 3).
FT CONFLICT 2687 2688 SV -> RL (in REF. 2).
SQ SEQUENCE 2871 AA; 331771 MW; 5639557CD469057 CRC64;

Query Match
Best Local Similarity 6.6%; Score 100; DB 1; Length 2871;
Matches 60; Conservative 57; Mismatches 93; Indels 128; Gaps 12;

QY 25 DLYNKYLDQVWPWTFDETIKELSRKQYSEASVAVG-----DIKVLMDSDQKYF 77
DB 1205 NLRKYTEINTKT---TKRISWCKDDSKNLRNQDRSRNRDLKQVNRINDSIL 1261
QY 78 EATQ-----TVYEWGVVTVLLSAYILLFDEYNEKQASQ---KDLIRILDDGVK 125
DB 1262 QATEQRRRAEENALQKACG-----SEIMQKQHLIELKQVNRQSRSEDNR 1308
QY 126 ---KLNRAQSKLLTSQSF-----NNASGKLLADS----- 153
DB 1309 HKQSLSEAAKTQDKNKEIERLKAEQFAKRRWYENELSKVRNNYDEEIIISLNQPET 1368
QY 154 -----QLTNDFSEKSSYFQSQVDRIKAEYAGAAAGIVAGPFGILLIISYTAAGVI 203

DB 1369 EINITKTTHLTQMKEEDTSGYRAQIDNLTENES----- 1404
QY 204 EGKLPENLRKTVONFFTSLSATVKQAKDIDAKL-----KLATEIAIGBI 253
DB 1405 -----LSEIKRLKNTLTQTTNLRVEEDIOQKATGSEVSKQOQLEVLQVQTM 1457
QY 254 KTEETTRFVVDYDMLSLKGAKKMINTNEYQOR 291
DB 1458 RTE-ESVRYKQSLDD-----AAKTIQDKNKEIER 1485

RESULT 44
MYH9 CHICK
ID MYH9 CHICK STANDARD; PRT; 1959 AA.
AC P14105;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, nonmuscle (Cellular myosin heavy chain) (NMHC).
GN MYH9.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestinal epithelium;
RX MEDLINE=90046668; PubMed=2813355;
RA Shohet R.V., Conti M.A., Kawamoto S., Preston Y.A., Brill D.A.,
RA Adelstein R.S.;
RT "Cloning of the cDNA encoding the myosin heavy chain of a vertebrate
cellular myosin."
RL Proc. Natl. Acad. Sci. U.S.A. 86:7726-7730(1989).
CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CAPING.
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 24-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; M26510; AAA48974.1; -.
DR PIR; A33977; A33977.
DR HSP; P10587; IBR2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin head.
DR InterPro; IPR004009; myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSTINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family.

FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
 FT DOMAIN 779 808 IQ.
 FT DOMAIN 837 1925 COILED COIL (POTENTIAL).
 FT NP_BIND 174 181 ATP (POTENTIAL).
 FT DOMAIN 654 676 ACTIN-BINDING.
 FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1959 AA; 226502 MW; A75C86086PD3A1A1 CRC64;

Query Match 6.6%; Score 99.5; DB 1; Length 1959;
 Best Local Similarity 20.1%; Pred.No.74; Indels 85; Gaps 9;
 Matches 62; Conservative 41; Mismatches 120; Indels 85; Gaps 9;

QY 7 EQTVVKAIE-----TADGALDLYNKYLDQVWPMTFDETIKE-----LSRFKQYS 55
 DB 1166 EQEVTYKTTDEAKTHEAQIQEMRQHSQAI--EELAEQLEQTKRVKANLEKAKQALE 1223
 QY 56 QEASVLGDIKVLMDSDQKYEATQTVVGVVTVLLSAYILLFDYNEKASAKDI 115
 DB 1224 SERAELSNVKKVLLQKQGA-----BHKRKKVDAQLOE 1256
 QY 116 LIRILDGKVKLENAQKSLTSSQSFNNASGKLLALDS---QLTDFSEKSYFOSQVDR 172
 DB 1257 LQVKFTEGERVKTLEAERVKNQLQVELDNTGLNQSDSKSIKAKDFSALESQLODTQEL 1316
 QY 173 IRKEAYAGAAAGIVAGPFGLIISYSIAAGVIEGK---LIPELNNKLTQVNFSTLSATV 229
 DB 1317 LQET-----RLKLSFSTKLQKQTEDEKNAKQLEBEEBAKRNLEKQISVLQ 1363
 QY 230 KQA-----NKDIDAAKLKATEIAAIGEIKTETTRFYVD 265
 DB 1364 QQAVARCKMDGLGCLGTAEBEAKKQLQKLSLQRYEIKIAYD--KLETKTQLQOE 1421
 QY 266 YDDMLSL 273
 DB 1422 LDIAVDL 1429

RESULT 45
 MYHA RAT
 ID MYHA RAT STANDARD; PRT; 1976 AA.
 AC QJUL70;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain, type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
 DE MYH10.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 EX MEDLINE=20483650; PubMed=11027611;
 RA Yam J.W.P., Chan K.W., Li N., Heiao W.L.W.;
 RT "Molecular cloning and functional analysis of the promoter region of rat nonmuscle myosin heavy chain-B gene";
 RL Biochem. Biophys. Res. Commun. 276:1203-1209(2000).
 CC -I- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS, CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND CAPPING.
 CC -I- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -I- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -I- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -I- SIMILARITY: Contains 1 IQ domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF139055; AAF61445.1; -;
 DR HSP; P10587; IBR2.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR001609; Myosin_head.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00663; myosin_head; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 KW Myosin: ATP-binding; Calmodulin-binding; Actin-binding;
 KW Coiled coil; Alkylation; Multigene family;
 FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
 FT DOMAIN 786 815 IQ.
 FT NP_BIND 178 185 COILED COIL (POTENTIAL).
 FT MOD_RES 701 701 ATP (POTENTIAL).
 FT MOD_RES 711 711 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1976 AA; 228963 MW; E32708BF9BF2B470 CRC64;

Query Match 6.6%; Score 99.5; DB 1; Length 1976;
 Best Local Similarity 21.5%; Pred.No.75; Indels 79; Gaps 10;
 Matches 60; Conservative 43; Mismatches 97; Indels 79; Gaps 10;

QY 10 VEVVKAIEATDAGALDLYNKYLDQV-----IPWKTETETIKEL-----SRFKQYSQ 57
 DB 1222 LERKNGLETDNKEACEVKKVQVKAESHEKELDAQVQELHAKVSEGLRVELAEK 1281
 QY 58 ASVL---VGDIKVLMDSDQKYEATQTVVGVVTVLLSAYILLFDYNEKASAKQD 114
 DB 1282 ANKLQNELDNTVLLPEAEKKGKMKFAK---DAAGLSQLODTQELQETQKLNLSR- 1337
 QY 115 ILIRILDGKVKLENAQKSLTSSQSFNNASGKLLALDSQLTDFSEKSYFOSQVDRIR 174
 DB 1338 --IRQLEEKNSLQEQEE---EEBAKRNLEKQVLAQLQSLADT-----KKVDD-- 1382
 QY 175 KEAYAGAAAGIVAGPFGLIISYSIAAGVIEGKLIPELNNKLTQVNFSTLSATVQANK 234
 DB 1383 -----DLGTIEG-----LEBAKKKLLK 1399
 QY 235 DIDAAKLKATEIAAIGEIKTETTRFYVDYDDMLSL 273
 DB 1400 DVEALSORLEEKVLAYD--KLETKTQLQELDDLTVDL 1436

RESULT 46
 RBP1 PLAVB
 ID RBP1 PLAVB STANDARD; PRT; 2869 AA.
 AC Q00798;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Reticulocyte binding protein 1 precursor.
 GN RBP1.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;

RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites"; 1226(1992).
RL Cell 63:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: Homodimer (Potential).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M8097; RAA29743.1; -
DR Malaria; Receptor; Signal; Transmembrane.
KW SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;

Query Match 6.6%; Score 99.5; DB 1; Length 2869;
Best Local Similarity 18.9%; Pred. No. 1.2e+02;
Matches 69; Conservative 71; Mismatches 132; Indels 93; Gaps 18;

Qy 10 VEVVKSALET--ADGALDLYNKYLDQVLPKTP--DE-----TIKELSPKQEQS 56
Db 1925 VSIYKVKESKHADYRRDANSYMSVMTLYANFLSDSEAKISSGWEFNAEKSFKTDLEL 1984

Qy 57 EASVLVGD-----IKVLMDSD-----KYFEATQTVYEWGVVTVLLSAYILLFDEYNEK-- 107
Db 1985 EIFSVINSNELAKKIQSDNDVTKERSEQLAKD-----ATDYNV-IKLKEFNEKLE 2039

Qy 108 KSAQKQILRLDGVKKLNEAQ-----KSLTSSQSFNNASGKLALDSQLTND 158
Db 2040 EAKKEEYVSEKREALKRLSQVREGIRCFHFENFRLLDNTLEENL-----KKQVVI 2091

Qy 159 FSEKSSYFOSQVDVIRKE--AYAGAAAGIVAGPPGLIISYSIAAGVTEGKLIPELNRLK 216
Db 2092 YRDKKSERESQLQWENEMNTYSNITQLE-----GIUVSAGESKEDIE--KLRSNEEMR 2145

Qy 217 TVQNFFTSLSATVQKANKID-----AAKLKLAETAAIGETKTET 257
Db 2146 NISEKISTIDSKVIEMNSTIDELYKLGKNCQAHWISLISYTNANKTSKLMINK---EK 2202

Qy 258 ETRFYVDY-----DDLMLSLKLG-----AAKMINTCNEYQ---QRHGKKTIF 298
Db 2203 ENTEKVDYIKDSSSDIGYVETLKGYGKSLTFSSASEIVQADTYSVNFPAKHESLN 2262

Qy 299 EPPDV 303
Db 2263 AIRDI 2267

RESULT 47

MSPI PLAFAC
ID MSPI PLAFAC STANDARD; PRT; 1726 AA.
AC P04934;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

NCBI TaxID=5035;
[1]
RX SEQUENCE OF 1-1103 FROM N.A.
RP MEDLINE=86205236; PubMed=3517809;
RA Weber J.L., Leininger W.M., Lyon J.A.;
RT "Variation in the gene encoding a major merozoite surface antigen of
the human malaria parasite Plasmodium falciparum";
RL Nucleic Acids Res. 14:3311-3323(1986).
RN [2]
CC
CC SEQUENCE OF 1104-1726 FROM N.A.
RP MEDLINE=88143999; PubMed=3278296;
RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
RT "Merozoite surface protein sequence from the Camp strain of the human
malaria parasite Plasmodium falciparum";
RL Nucleic Acids Res. 16:1206-1206(1988).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(Potential).
CC
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X03831; CAA27446.1; -
DR PIR; A23386; SAZQGM.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196197 MW; DDBAD45FA352BCF3 CRC64;

Query Match 6.5%; Score 99; DB 1; Length 1726;
Best Local Similarity 18.6%; Pred. No. 69;
Matches 56; Conservative 54; Mismatches 99; Indels 92; Gaps 13;

Qy 28 NKYLQVTPWKTFTDIKE-LSRFKQEQYSQ-----EASVLVGDIKVL- 68
Db 1322 DEYLDQVVTGEAISVTMDNLSGPNEDYVILKPLAGVYRSKQIEKNIETFNLD 1381

Qy 69 LMDs---QDKYFEATQTVYEWGVVTVLLSAYILLFDEYNEKKSASAKDLIRILDGVK 125
Db 1382 ILNRLKRRKYP-----LDVLESDLVQFRRHSNE-----YIEDSFK 1419

Qy 126 KLNEAQKSLTSSQSFNNASGKLALDSQLTNDF---SEKSSYFQ-----SQ 169
Db 1420 LLNSEQKNTLLSKYKY-----IKESVENDIKFAQEGISYVEKVLAKYKDLIESIKK 1470

Qy 170 VDRIRKEAVAGAAAGIVAGPPGLIISYSIAAGVTEGKLIPEL-----NNELKTVQNF 222
Db 1471 VIKERKEKFPSPPTTPSP-----AKTDQKESKELPFLTIETIYNNLVNKIDYLL 1524

Qy 223 TSLSATVQKANKIDDAKLKLAETAAIGETKTETTRFYVDYDMLSLKGAAGKVI 282

1525 INLRKINDCNVKEAHVKI-TKLSDLKATDDKIDLFKNHDFE-----AIKKLI 1574

283 N 283

1575 N 1575

RESULT 48

MACF_HUMAN
ID MACF_HUMAN STANDARD; PRT: 5430 AA.
AC Q9UPN3; Q75053; Q9WXY2; Q9H540; Q9UKP0; Q9ULG9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Microtubule-actin crosslinking factor 1, isoforms 1/2/3 (Actin cross-
linking family protein 7) (Macrophilin 1) (Trabeculin-alpha) (620 kDa
actin-binding protein) (ABP620).
GN MACF1 OR ACF7 OR ABP620 OR KIAA0465 OR KIAA1251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=20001959; PubMed=10529403;
RA Okuda T., Matsuda S., Nakatsugawa S., Ichigotani Y., Iwahashi N.,
Takahashi M., Ishigaki T., Hamaguchi M.;
RT "Molecular cloning of macrophilin, a human homologue of Drosophila
kakapo with a close structural similarity to plectin and dystrophin.";
RL Biochem. Biophys. Res. Commun. 264:568-574 (1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20026884; PubMed=10559237;
RA Sun Y., Zhang J., Kraeft S.-K., Auclair D., Chang M.-S., Liu Y.,
Sutherland R., Salgia R., Griffin J.D., Ferland L.H., Chen L.B.;
RT "Molecular cloning and characterization of human trabeculin-alpha, a
giant protein defining a new family of actin-binding proteins.";
RL J. Biol. Chem. 274:33522-33530 (1999).
RN [3]
RP SEQUENCE OF 182-4812 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=21833812; PubMed=11845288;
RA Gong T.-W.L., Besirli C.G., Lonax M.I.;
RT "MACF1 gene structure: a hybrid of plectin and dystrophin.";
RL Mamm. Genome 12:852-861 (2001).
RN [4]
RP SEQUENCE OF 868-2350 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 6:337-345 (1999).
RN [5]
RP SEQUENCE OF 1544-5057 FROM N.A.
RA Corby N.;
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 3312-5430 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106 (2002).
RN [7]
RP SEQUENCE OF 3734-5430 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
Nakajima D., Nomura N., Ohara O.;

RT "Characterization of cDNA clones in size-fractionated cDNA libraries
from human brain.";
RL DNA Res. 4:345-349 (1997).
CC -|- FUNCTION: F-ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-
LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO
CC MICROTUBULES (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=2;
CC IsoId=Q9UPN3-2; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9UPN3-1; Sequence=VSP_007341;
CC Name=3;
CC IsoId=Q9UPN3-3; Sequence=Not described;
CC Name=4;
CC IsoId=Q9PK2-1; Sequence=External;
CC -|- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -|- SIMILARITY: Belongs to the plectin or cytolinker family.
CC -|- SIMILARITY: Contains 1 actin-binding domain.
CC -|- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -|- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -|- SIMILARITY: Contains 1 SH3 domain.
CC -|- SIMILARITY: Contains 37 spectrin repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB029290; BAA83821.1; -;
CC EMBL; AF141368; AAF06360.1; -;
CC EMBL; AF325341; AAL39000.1; -;
CC EMBL; AF325330; AAL39000.1; JOINED.
CC EMBL; AF325331; AAL39000.1; JOINED.
CC EMBL; AF325332; AAL39000.1; JOINED.
CC EMBL; AF325333; AAL39000.1; JOINED.
CC EMBL; AF325334; AAL39000.1; JOINED.
CC EMBL; AF325335; AAL39000.1; JOINED.
CC EMBL; AF325336; AAL39000.1; JOINED.
CC EMBL; AF325339; AAL39000.1; JOINED.
CC EMBL; AF325340; AAL39000.1; JOINED.
CC EMBL; AB033077; BAA86565.1; -;
CC EMBL; AL137853; CAC15920.1; -;
CC EMBL; AB007934; BAA32310.2; -;
CC PIR; T00079; T00079.
CC HSSP; Q01082; IBKR.
CC Genew; HGNC:13664; MACF1.
CC GO; GO:0005856; C:cytoskeleton; NAS.
CC GO; GO:0003780; F:actin cross-linking activity; NAS.
CC GO; GO:0005509; F:calcium ion binding activity; NAS.
CC GO; GO:0008017; F:microtubule binding activity; NAS.
CC InterPro; IPR001589; Actbind_actnin.
CC InterPro; IPR001715; Calponin-like.
CC InterPro; IPR003108; GAS2.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR001452; SH3.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF00307; CH; 2.
CC Pfam; PF00036; ehand; 2.
CC Pfam; PF02187; GAS2; 1.
CC Pfam; PF00435; spectrin; 27.
CC ProDom; PD000012; EF-hand; 1.
CC SMART; SM00033; CH; 2.
CC SMART; SM00054; EFh; 2.
CC SMART; SM00243; GAS2; 1.
CC SMART; SM00150; SPEC; 36.
CC PROSITE; PS00019; ACTININ_1; 1.
CC PROSITE; PS00020; ACTININ_2; FALSE_NEG.
CC PROSITE; PS50021; CH; 2.

DR PROSITE; PS00018; EF HAND; 2.
 DR PROSITE; PS0002; SH3; FALSE NEG.
 KW Actin-binding; Cytoskeleton; Calcium; Calcium-binding; Repeat;
 KW SH3 domain; Coiled coil; Alternative splicing.
 FT DOMAIN 1 295
 FT DOMAIN 78 181
 FT DOMAIN 194 295
 FT DOMAIN 243 265
 FT DOMAIN 477 529
 FT DOMAIN 722 751
 FT DOMAIN 816 843
 FT DOMAIN 1013 1118
 FT DOMAIN 1164 1191
 FT DOMAIN 1399 1690
 FT DOMAIN 1780 1843
 FT DOMAIN 1975 2005
 FT DOMAIN 2039 2312
 FT DOMAIN 2385 2417
 FT DOMAIN 2544 2695
 FT DOMAIN 2760 2838
 FT DOMAIN 2911 3001
 FT DOMAIN 3130 3164
 FT DOMAIN 3244 3277
 FT DOMAIN 3418 3482
 FT DOMAIN 3596 3666
 FT DOMAIN 3786 3806
 FT DOMAIN 3852 3931
 FT DOMAIN 3967 3987
 FT DOMAIN 4084 4218
 FT DOMAIN 4343 4378
 FT DOMAIN 4408 4437
 FT DOMAIN 4468 4498
 FT DOMAIN 4507 4935
 FT DOMAIN 5044 5067
 FT REPEAT 314 355
 FT REPEAT 591 623
 FT REPEAT 680 784
 FT REPEAT 786 800
 FT REPEAT 871 923
 FT REPEAT 1250 1272
 FT REPEAT 1287 1342
 FT REPEAT 1455 1534
 FT REPEAT 1547 1659
 FT REPEAT 1815 1891
 FT REPEAT 1932 2042
 FT REPEAT 2260 2280
 FT REPEAT 2372 2395
 FT REPEAT 2398 2507
 FT REPEAT 2510 2618
 FT REPEAT 2621 2728
 FT REPEAT 2731 2838
 FT REPEAT 2841 2945
 FT REPEAT 2987 3024
 FT REPEAT 3136 3163
 FT REPEAT 3187 3274
 FT REPEAT 3277 3383
 FT REPEAT 3386 3492
 FT REPEAT 3495 3601
 FT REPEAT 3604 3673
 FT REPEAT 3713 3819
 FT REPEAT 3832 3927
 FT REPEAT 3982 4043
 FT REPEAT 4046 4152
 FT REPEAT 4155 4262

Query Match
 Best Local Similarity 6.5%; Score 98.5; DB 1; Length 5430;
 Matches 53; Conservative 48; Mismatches 113; Indels 27; Gaps 8;
 29 KYLDQVIVPKVTFDETIKELSRFKQYEQASVVLGDIKVLMDSDQKYFATQTVYWGCG 88
 1404 KYISDAL--RRLEEEKVVEEKQEHVEKVELLGVSTLARNYQK--ATISSETKEST 1458

89 VVTOLLSAYILLFDEYNEK--ASAOKDILIRLDGVKKLNEAOKSLTSSOFNNA- 144
 1459 DIEKALEQQVLSBELTTTKQVSEAIKASQIFLAKHGKLSKEKKQISQLNALNKAY 1518
 145 -----SGKLLALDSQLTNDPFSKSYFQSQVDRIRKE-----AYAGAAAGIVAGPGL 192
 1519 HDLCDSGANQLQQLSQLAHQTEQK--LQKQNTCHQQLBDCGWSVYQAEARALAGHQR 1576
 193 IISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLGATVQKANKDIDAAKLKLA-TEIAAIG 251
 1577 TQOQLSA---LQKQSDLKLODDIQRATSPATVYKQIEGFMEENQTKLSPRELTALR 1633
 252 E 252
 1634 E 1634

RESULT 49
 SNEI_HUMAN
 ID SNEI_HUMAN STANDARD; PRT; 8797 AA.
 AC Q8NF91; Q94890; Q8N9P7; Q8TCPI; Q8WMW6; Q8WXXF6; Q96N17;
 AC Q9COA7; Q9H525; Q9H526; Q9NS36; Q9NU50; Q9UJ06; Q9UJ07; Q9ULF8;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Nesprin 1 (Nuclear envelope spectrin repeat protein 1) (Synaptic
 DE nuclear envelope protein 1) (Syn-1) (Myocyte nuclear envelope protein
 DE 1) (Myne-1) (Enaptin).
 GN SNEI OR MYNE1 OR KIAA0796 OR KIAA1756 OR KIAA1262.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), CHARACTERIZATION, AND
 RP MUTAGENESIS OF 8758-LEU-CYS-8763.
 RC TISSUE=Heart, Placenta, Skeletal muscle, Spleen, and Testis;
 RX MEDLINE=21652858; PubMed=11792814;
 RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,
 RA Weissberg P.L., Ellis J.A., Shanahan C.M.;
 RT "Nesprins: a novel family of spectrin-repeat-containing proteins that
 RT localize to the nuclear membrane in multiple tissues.";
 RL J. Cell Sci. 114:4485-4498(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND SUBCELLULAR LOCATION.
 RC TISSUE=Heart, Spleen, and Testis;
 RX MEDLINE=22396983; PubMed=12408964;
 RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;
 RT "The nesprins are giant actin-binding proteins, orthologous to
 RT Drosophila melanogaster muscle protein MSP-300.";
 RL Genomics 80:473-481(2002).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 4), AND VARIANT GLY-8323.
 RA Braune S., Abraham S., Padmakumar V., Tunggal B., Noegel A.A.,
 RA Korenbaum E.;
 RT "The longest isoform of enaptin/Syne-1, a nuclear envelope associated
 RT protein, binds actin cytoskeleton via the alpha-actinin-like actin-
 RT binding domain.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 8 AND 9).
 RA Zhang Q., Shanahan C.M.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Almeida J., Clark S., Griffiths C., Lloyd D., Parker A., Smith M.,
 RA Tracey A., Williams S.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 1-856 FROM N.A.
 RC TISSUE=Kidney;
 RA Gough L., Fan J., Lisa G., Chu S., Winnick S., Beck K.A.;
 RT "Golgi localization of syne-1.";

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 28-778 AND 2901-3476 FROM N.A.
RC TISSUE=Adrenal gland, and Teratocarcinoma;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yanazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Takenoto M., Ota T., Suzuki Y., Sugano S., Nagahari K., Masuho Y.,
RA Nagai K., Isogai T.;
RT "NED0 human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 443-8797 FROM N.A. (ISOFORM 5).
RC TISSUE=Brain;
RX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
RN [9]
RP SEQUENCE OF 743-8797 FROM N.A. (ISOFORM 6).
RC TISSUE=Brain;
RA Ansoerge W., Wirkner U., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE OF 4219-8797 FROM N.A. (ISOFORM 7).
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. xv.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
RN [11]
RP SEQUENCE OF 6922-8797 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [12]
RP REVISIONS.
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [13]
RP SEQUENCE OF 7631-8797 FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=21659781; PubMed=11801724;
RA Maslow J.M.K., Kim M.S., Davis D.B., McNally E.M.;
RT "Myne-1, a spectrin repeat transmembrane protein of the myocyte inner
RT nuclear membrane, interacts with lamin A/C.";
RL J. Cell Sci. 115:61-70(2002).
RN [14]
RP SEQUENCE OF 8406-8797 FROM N.A.
RA Ma F.-R., Zhu L.-P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Involved in the maintenance of nuclear organization and
CC structural integrity. Probable anchoring protein which tethers the
CC nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
CC by interacting with the nuclear envelope and with F-actin in the
CC cytoplasm.

CC -1- SUBUNIT: Interacts with MUSK, with F-actin via its N-terminal
CC domain, and with LMNA in vitro (By similarity).
CC -1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
CC largest part of the protein is cytoplasmic, while its C-terminal
CC part is associated with the nuclear envelope, most probably the
CC outer nuclear membrane. In skeletal and smooth muscles, a
CC significant amount is found in the sarcomeres.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=9;
CC Name=1;
CC IsoId=Q8NF91-1; Sequence=Displayed;
CC Name=2; Synonyms=Beta;
CC IsoId=Q8NF91-2; Sequence=VSP_007130;
CC Name=3; Synonyms=Alpha;
CC IsoId=Q8NF91-3; Sequence=VSP_007132, VSP_007144;
CC Name=4;
CC IsoId=Q8NF91-4; Sequence=VSP_007134, VSP_007139, VSP_007140,
CC VSP_007144;
CC Name=5;
CC IsoId=Q8NF91-5; Sequence=VSP_007135, VSP_007136;
CC Note=No experimental confirmation available;
CC Name=6;
CC IsoId=Q8NF91-6; Sequence=VSP_007137, VSP_007138;
CC Note=No experimental confirmation available;
CC Name=7;
CC IsoId=Q8NF91-7; Sequence=VSP_007141, VSP_007142;
CC Note=No experimental confirmation available;
CC Name=8; Synonyms=Beta 2;
CC IsoId=Q8NF91-8; Sequence=VSP_007131;
CC Name=9; Synonyms=Alpha 2;
CC IsoId=Q8NF91-9; Sequence=VSP_007133, VSP_007143, VSP_007144;
CC -1- TISSUE SPECIFICITY: Widely expressed. Highly expressed in skeletal
CC and smooth muscles, heart, spleen, and peripheral blood
CC leukocytes.
CC -1- DOMAIN: The Klarsicht domain, which contains a transmembrane
CC domain, mediates the nuclear envelope targeting.
CC -1- SIMILARITY: Belongs to the Nesprin family.
CC -1- SIMILARITY: Contains 1 actin-binding domain.
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -1- SIMILARITY: Contains 12 HAT repeats.
CC -1- SIMILARITY: Contains 1 Klarsicht domain.
CC -1- SIMILARITY: Contains 31 spectrin repeats.
CC -1- CAUTION: Ref.5 (CAB55865, CAB55866, CAC16280 and CAC16281)
CC sequences differ from that shown due to erroneous gene model
CC prediction.
CC -1- CAUTION: Ref.7 (BAB71097) sequence differs from that shown due to
CC a chimeric cDNA.
CC -1- CAUTION: Ref.14 sequence differs from that shown due to two
CC frameshifts in positions 8412 and 8784.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY061755; AAL33798.1; -
DR EMBL; AY061756; AAL33799.1; -
DR EMBL; AF495910; AAN60442.1; -
DR EMBL; AF535142; AAK03486.1; -
DR EMBL; AY184203; AAO27771.1; -
DR EMBL; AY184206; AAO27774.1; -
DR EMBL; AL049548; CAB55865.1; ALT_SEQ.
DR EMBL; AL049548; CAB55866.1; -
DR EMBL; AL078582; CAB87586.1; -
DR EMBL; AL136079; -; NOT_ANNOTATED_CDS.
DR EMBL; AL136832; CAC16280.1; ALT_SEQ.
DR EMBL; AL138832; CAC16281.1; ALT_SEQ.
DR EMBL; AL1357081; -; NOT_ANNOTATED_CDS.
DR EMBL; AL450401; -; NOT_ANNOTATED_CDS.
DR EMBL; AL589963; -; NOT_ANNOTATED_CDS.

Query Match 6.5%; Score 98.5; DB 1; Length 8797;
Best Local Similarity 22.0%; Pred. No. 5.6e+02;
Matches 67; Conservative 44; Mismatches 123; Indels 71; Gaps 14;

QY 8 QTVVWKSALETADGALDLY-----NKYLDQVPMKTFDETIKELSRFKQYSEASV 60
DB 2283 QKLEHAKETEVAKGLTKDTAQSTQVEKINDITW--FTKVEESLMNCAQNETCEALK 2340
QY 61 LVGDIKVLMDSDQKYFEATQTVYVWGVVTVLLSAILLFDYNEKASAKDILIRIL 120
DB 2341 KVXDIQKELQSQSNISSTOENLNSLC-----RYHSALLESIGRAM 2382
QY 121 DDGKKLNEAKSKSLTSSQSFNNASGKLLALDSLTNDFFSEKSYFQSDVRIRKAYAG 180
DB 2383 TGLIKK-HEAVSOLCSKTOA-----SLOESLEKHFSESQEFQ-----EWF LG 2424
QY 181 AAGIVAGPGLIISYIAG--VIEGKLIPELNNELKTVQNFSTLSATVQANKDID 237
DB 2425 AKAAAKES-----SDRTGDSKVLEAKL-HDLQNLDSVDSGQSKLDA-VTQEGQTL 2474
QY 238 AAKKLKATEAAGEIKETETTFYVYDDMLSLKGAACKMINTC---NEYQORHG 293
DB 2475 A--HLSKQIVS-----SIOEQIKANEFPQAFKQCLKD--KQALQDCASELGSFPDQHR 2525
QY 294 KKTLP 298
DB 2526 KLNW 2530

RESULT 50
RASO_THEME
ID RASO_THEME STANDARD; PRT; 852 AA.
AC Q9X1X1;
RC STRAIN=MSB8 / DSM 3109;
EX MEDLINE=93287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima."
CC Nature 399:323-329(1999).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ch/announcement/>
CC or send an email to license@ebi.ebi.ch).
CC -----
DB EMBL; AE001806; AAD36703.1; -

DR PIR: D72230; D72230.
DR TIGR; TM1636; -;
DR HAMAP; MF_00445; -; 1.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PFO2463; SMC_N; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 155 711 COILED COIL (POTENTIAL).
SQ SEQUENCE 852 AA; 100001 MW; 31BA9F72A4EC5CD2 CRC64;
Query Match 6.5%; Score 98; DB 1; Length 852;
Best Local Similarity 17.1%; Pred. No. 33;
Matches 56; Conservative 69; Mismatches 124; Indels 78; Gaps 11;

QY 7 EQTVVWKSALETADGALDLYNKYLDQVPMKTFDETIKELSRFKQY----- 54
DB 334 BEIGEKIKEEDLLLEKVLKASRPILLEQRILK-----ENLRLLEEFRLVGEKEKEK 387
QY 55 -----SQEASVLVGDIKVLLMDSQDKY--FEATQTVYEW-----CGVV---T 91
DB 388 ELLSTIKETENETKNELEKLDDELGIKLDHM-KWLAVQIASSLNEGDTCPVCGGVFHGV 446
QY 92 QLLSAVILLFDYNEKASAKQ-----KDLIRILDDGVKKLNEAKSKSLTSSQSFN 142
DB 447 EAVENIDFEKLDQKRSLENTLVKSKSLSLIEDLLMKIEGKKLKSIRNQIE 506
QY 143 NASGKLALDSLTNDFFSEKSYFQSDVRIRKAYAGAAAGIVAGPGLIISYIAAGV 202
DB 507 KIEEELHRLG--YSEDLEEKLDKRRKLRKIEERHS-----ISQKITAAD 550
QY 203 IEGKLIPELNNRLKTVQNFSTLSATVQANKDIDAKKLKATEIAATG-----EIK 254
DB 551 VQ---ISQENQLKEIKGEIAKRETLKEQRENDQLKSDFFDRKIGIGFEFRILVK 607
QY 255 TETETTFRYVDYDDMLSLKGAACKM 281
DB 608 EVKDAEKELGVVETEIRLLESLEL 634

Search completed: November 28, 2003, 13:49:45
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2003, 13:48:16 ; Search time 41 seconds
(without alignments)

1919.660 Million cell updates/sec

Title: US-09-993-292A-2

Perfect score: 1515

Sequence: 1 MTSIFARQTVVVKSAIETA.....NEYQCRHGKTLPEVDVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 259052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL 23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_oxganelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1503	99.2	303	Q934C4	Q934C4 salmonella
2	1503	99.2	305	Q8Z727	Q8Z727 salmonella
3	1466	95.8	303	Q83RR6	Q83RR6 salmonella
4	1393	91.9	303	Q8VU70	Q8VU70 escherichia
5	1387	91.6	305	Q8REB3	Q8REB3 escherichia
6	1110	73.3	300	Q9X2S8	Q9X2S8 escherichia
7	519	34.3	113	Q9RMG1	Q9RMG1 shigella fl
8	519	34.3	115	Q9RCT3	Q9RCT3 shigella fl
9	395	26.1	93	Q8FI27	Q8FI27 escherichia
10	130.5	8.6	895	Q9LIW7	Q9LIW7 oryza sativ
11	122	8.1	495	P71497	P71497 mycoplasma
12	118	7.8	1496	Q9SZK7	Q9SZK7 arabidopsis
13	115	7.6	478	P79WH8	P79WH8 sulfolobus
14	114.5	7.6	478	Q64067	Q64067 bacterioph
15	114.5	7.6	478	Q31954	Q31954 bacillus su
16	114.5	7.6	588	Q9VGA8	Q9VGA8 drosophila

17	114.5	7.6	652	2	Q9X360	Q9X360 bacillus an
18	114	7.5	720	4	Q8H6Q7	Q8H6Q7 homo sapien
19	114	7.5	1189	16	Q8CPI0	Q8CPI0 staphylococ
20	113.5	7.5	538	2	Q9Z429	Q9Z429 pseudomonas
21	113.5	7.5	727	16	Q8ZON2	Q8ZON2 anabaena sp
22	113.5	7.5	1127	12	Q9YV76	Q9YV76 melanoplus
23	113.5	7.5	3102	5	Q45614	Q45614 caenorhabdi
24	111.5	7.4	1999	11	Q63731	Q63731 rattus norv
25	111.5	7.4	2819	16	Q98Q88	Q98Q88 mycoplasma
26	111	7.3	1295	5	Q22257	Q22257 caenorhabdi
27	110.5	7.3	350	16	Q8XNA0	Q8XNA0 clostridium
28	110.5	7.3	504	5	Q81752	Q81752 plasmodium
29	110.5	7.3	649	10	Q9LVQ4	Q9LVQ4 arabidopsis
30	110.5	7.3	828	16	Q8EWP9	Q8EWP9 mycoplasma
31	110.5	7.3	927	16	Q92882	Q92882 listeria in
32	110.5	7.3	979	11	Q3EP71	Q3EP71 mus musculu
33	110	7.3	821	3	Q8581	Q8581 saccharomyc
34	109.5	7.2	852	5	Q81754	Q81754 plasmodium
35	109.5	7.2	1024	2	Q8GA40	Q8GA40 escherichia
36	109.5	7.2	1024	2	Q8G9Z4	Q8G9Z4 escherichia
37	109.5	7.2	1024	16	Q8FE01	Q8FE01 escherichia
38	109.5	7.2	1175	16	Q8XNW6	Q8XNW6 clostridium
39	109.5	7.2	1492	2	Q68724	Q68724 versinia pe
40	109.5	7.2	1545	16	Q9ZH03	Q9ZH03 versinia pe
41	109.5	7.2	2752	5	Q9B5Y0	Q9B5Y0 plasmodium
42	109	7.2	595	17	Q9UYA4	Q9UYA4 pyrococcus
43	109	7.2	739	17	Q9UYF8	Q9UYF8 pyrococcus
44	109	7.2	1098	16	Q50733	Q50733 borrelia bu
45	108.5	7.2	296	2	Q933C4	Q933C4 escherichia
46	108.5	7.2	792	10	Q9AR09	Q9AR09 oryza sativ
47	108.5	7.2	3692	16	Q8CNU9	Q8CNU9 staphylococ
48	108	7.1	403	2	Q84555	Q84555 staphylococ
49	108	7.1	549	16	Q8R8D8	Q8R8D8 thermocact
50	108	7.1	726	2	Q9RMD9	Q9RMD9 acinetobact
51	108	7.1	998	2	Q46716	Q46716 escherichia
52	108	7.1	998	2	P71223	P71223 escherichia
53	108	7.1	2139	5	Q07569	Q07569 entamoeba h
54	107.5	7.1	630	5	Q25613	Q25613 onchocerca
55	107	7.1	851	16	Q8REB7	Q8REB7 fusobacteri
56	107	7.1	955	2	Q9EUD4	Q9EUD4 pasteurella
57	107	7.1	955	2	Q9EV22	Q9EV22 pasteurella
58	107	7.1	998	2	Q85101	Q85101 escherichia
59	107	7.1	1312	11	Q9JII8	Q9JII8 rattus norv
60	106.5	7.0	539	16	Q9X0N0	Q9X0N0 thermotoga
61	106.5	7.0	927	16	Q8Y432	Q8Y432 listeria mo
62	106.5	7.0	1173	4	Q9UP86	Q9UP86 homo sapien
63	106.5	7.0	1272	16	Q98PR8	Q98PR8 mycoplasma
64	106.5	7.0	1312	4	Q92878	Q92878 homo sapien
65	106.5	7.0	1316	10	Q9SL02	Q9SL02 arabidopsis
66	106.5	7.0	1499	2	Q54186	Q54186 streptococ
67	106.5	7.0	2346	5	Q9V624	Q9V624 drosophila
68	106.5	7.0	2346	5	Q01385	Q01385 drosophila
69	106.5	7.0	735	5	Q9VMT8	Q9VMT8 drosophila
70	106	7.0	685	5	Q9V6C3	Q9V6C3 drosophila
71	106	7.0	998	2	Q47262	Q47262 escherichia
72	106	7.0	998	2	Q47461	Q47461 escherichia
73	106	7.0	998	2	Q9LC58	Q9LC58 escherichia
74	106	7.0	1364	5	Q97036	Q97036 hydra magni
75	105.5	7.0	540	1	Q33398	Q33398 natronobact
76	105.5	7.0	540	1	Q93398	Q93398 natronobact
77	105.5	7.0	1273	5	Q95H02	Q95H02 caenorhabdi
78	105.5	7.0	1318	4	Q43254	Q43254 homo sapien
79	105	6.9	1575	2	Q54185	Q54185 streptococ
80	105	6.9	1601	16	Q92827	Q92827 listeria in
81	104.5	6.9	295	5	Q813F5	Q813F5 plasmodium
82	104.5	6.9	318	16	Q8E0D4	Q8E0D4 streptococ
83	104.5	6.9	566	3	Q12321	Q12321 saccharomyc
84	104.5	6.9	1051	16	Q8EMQ0	Q8EMQ0 mycoplasma
85	104.5	6.9	1429	10	Q9LXV6	Q9LXV6 arabidopsis
86	104.5	6.9	1430	5	Q9VMA7	Q9VMA7 drosophila
87	104	6.9	411	16	Q8D3E6	Q8D3E6 wiggleswort
88	104	6.9	633	3	Q59795	Q59795 schizosacch
89	104	6.9	809	16	Q97GR5	Q97GR5 clostridium

90 104 6.9 901 16 Q9CF11 Q9CF11 lactococcus
 91 104 6.9 1474 5 Q23870 Q23870 dictyosteli
 92 104 6.9 1569 2 Q54183 Q54183 streptococ
 93 104 6.9 1938 6 Q9GJP9 Q9GJP9 oryctolagus
 94 103.5 6.8 284 5 Q9GZ70 Q9GZ70 perna virid
 95 103.5 6.8 318 16 Q8E608 Q8E608 streptococ
 96 103.5 6.8 664 16 Q97FV9 Q97FV9 clostridium
 97 103.5 6.8 1583 4 Q15045 Q15045 homo sapien
 98 103.5 6.8 1964 13 Q93522 Q93522 xenopus lae
 99 103.5 6.8 2011 5 Q8MMC2 Q8MMC2 drosophila
 100 103.5 6.8 2017 5 Q94992 Q94992 drosophila

ALIGNMENTS

RESULT 1

Q934C4 PRELIMINARY; PRT; 303 AA.
 AC Q934C4;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Cytolysin A.
 GN CytA.
 OS Salmonella enterica subsp. enterica serovar Typhi, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=90370, 601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. enterica subsp. enterica serovar Typhi, and S. typhi;
 RC STRAIN=Ty21a, and SMI S2369/96;
 RA Oscarsson J., Westermarck M., Lofdahl S., Uhlin B.;
 RT "Expression of a pore-forming cytotoxin by Salmonella ser. Typhi and
 RT Salmonella ser. Paratyphi A";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. enterica subsp. enterica serovar Typhi, and S. typhi;
 RC STRAIN=Ty21a, and SMI S2369/96;
 RA Oscarsson J.;
 RL Thesis (1999), Umea University, Sweden.
 DR EMBL; AJ313034; CAC38363.1; -
 DR EMBL; AJ313032; CAC38360.1; -
 DR InterPro; IPR000209; Peptidase_s8.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 SQ SEQUENCE 303 AA; 33789 MW; 0D1B4AD566AB972C CRC64;

Query Match 99.2%; Score 1503; DB 2; Length 303;
 Best Local Similarity 99.7%; Pred. No. 4.1e-90;
 Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MTGIFAEQTVVEVVKSAIETADGALDLYNKYLDQVDPWKTDETIKLSRFKQYSQASV 60
 Db 1 MTGIFAEQTVVEVVKSAIETADGALDLYNKYLDQVDPWKTDETIKLSRFKQYSQASV 60
 Qy 61 LVGDIKVLMDSDQKYEATQTVYEWGVVTVQLLSAYILLDFDYEKKASAKDILIRIL 120
 Db 61 LVGDIKVLMDSDQKYEATQTVYEWGVVTVQLLSAYILLDFDYEKKASAKDILIRIL 120
 Qy 121 DDGVKVLNEAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSVQDRIRKEAYAG 180
 Db 121 DDGVKVLNEAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSVQDRIRKEAYAG 180
 Qy 181 AAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLATVKQANKDIDAAK 240
 Db 181 AAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLATVKQANKDIDAAK 240
 Qy 241 LKLAETAAIGEIKTETETTRFFVYDDMLSLKGAKKWINTCNEYQQRHGKTLFEV 300
 Db 241 LKLAETAAIGEIKTETETTRFFVYDDMLSLKGAKKWINTCNEYQQRHGKTLFEV 300

RESULT 3

Q93RR6 PRELIMINARY; PRT; 303 AA.
 ID Q93RR6
 AC Q93RR6;

Qy 301 PDV 303
 Db 301 PDV 303
 RESULT 2
 ID Q8Z727 PRELIMINARY; PRT; 305 AA.
 AC Q8Z727;
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Haemolysin HlyE.
 GN HlyE.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 EX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18";
 RL Nature 413:848-852(2001).
 DR EMBL; AL627270; CAD01758.1; -
 DR InterPro; IPR000209; Peptidase_S8.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Complete proteome.
 SQ SEQUENCE 305 AA; 34034 MW; EFCB7F86CB907B6C CRC64;

Query Match 99.2%; Score 1503; DB 16; Length 305;
 Best Local Similarity 99.7%; Pred. No. 4.1e-90;
 Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MTGIFAEQTVVEVVKSAIETADGALDLYNKYLDQVDPWKTDETIKLSRFKQYSQASV 60
 Db 3 MTGIFAEQTVVEVVKSAIETADGALDLYNKYLDQVDPWKTDETIKLSRFKQYSQASV 62
 Qy 61 LVGDIKVLMDSDQKYEATQTVYEWGVVTVQLLSAYILLDFDYEKKASAKDILIRIL 120
 Db 63 LVGDIKVLMDSDQKYEATQTVYEWGVVTVQLLSAYILLDFDYEKKASAKDILIRIL 122
 Qy 121 DDGVKVLNEAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSVQDRIRKEAYAG 180
 Db 123 DDGVKVLNEAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSVQDRIRKEAYAG 182
 Qy 181 AAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLATVKQANKDIDAAK 240
 Db 183 AAAGIVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLATVKQANKDIDAAK 242
 Qy 241 LKLAETAAIGEIKTETETTRFFVYDDMLSLKGAKKWINTCNEYQQRHGKTLFEV 300
 Db 243 LKLAETAAIGEIKTETETTRFFVYDDMLSLKGAKKWINTCNEYQQRHGKTLFEV 302
 Qy 301 PDV 303
 Db 303 PDV 305

DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Cytolysin A.
 GN CLVA.
 OS Salmonella paratyphi-a.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=54388;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SMI S3068/99;
 RA Oscarsson J., Westermarck M., Lofdahl S., Unlin B.;
 RT "Expression of a pore-forming cytotoxin by Salmonella ser. Typhi and
 RT Salmonella ser. Paratyphi A";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SMI S3068/99;
 RA Oscarsson J.;
 RL Thesis (1999), Department of Microbiology, Umea University, Umea,
 RL Sweden.
 DR EMBL; AJ313033; CAC38362.1; --
 DR InterPro; IPR000209; Peptidase S8.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 SQ SEQUENCE 303 AA; 33744 MW; 041F03F2018CCAE3 CRC64;

 Query Match 96.8%; Score 1466; DB 2; Length 303;
 Best Local Similarity 97.0%; Pred. NO. 1e-87; Indels 0; Gaps 0;
 Matches 294; Conservative 2; Mismatches 7;

 QY 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLQVLPWKTFDETIKELSRFKQVSEASV 60
 DB 1 MTGIFAEQTVVVKSAIETADGALDLYNKYLQVLPWKTFDETIKELSRFKQVSEASV 60

 QY 61 LVGDIKVLMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQKIDILIRIL 120
 DB 61 LVGDIKVLMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQKIDILIRIL 120

 QY 121 DDGVKLNKAEQSLTSSQSFNNASGKLLALDSQLTNDFSEKSYFQSOVDRIKREAYAG 180
 DB 121 DDGVKLNKAEQSLTSSQSFNNASGKLLALDSQLTNDFSEKSYFQSOVDRIKREAYAG 180

 QY 181 AAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQNFTSLSATVQANKDIDAAK 240
 DB 181 AAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQNFTSLSATVQANKDIDAAK 240

 QY 241 LKLAETIAAIGETETETTRFYVDYDMLSLKGAACKMINTCNEYQQRHGKKTLEFV 300
 DB 241 LKLAETIAAIGETETETTRFYVDYDMLSLKGAACKMINTCNEYQQRHGKKTLEFV 300

 RESULT 4
 QYVU70 ID Q8VU70 PRELIMINARY; PRT; 303 AA.
 AC Q8VU70;
 DT 01-MAR-2002 (TREMELrel. 20, Created)
 DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
 DE Cytolysin A.
 GN CLVA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH9802;
 RA Chang G.-N., Ho K.-C.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF240780; AAL55667.1; --
 DR InterPro; IPR000209; Peptidase S8.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 SQ SEQUENCE 303 AA; 33800 MW; 7A09C193E9287A6A CRC64;

 Query Match 91.9%; Score 1393; DB 2; Length 303;
 Best Local Similarity 90.4%; Pred. NO. 5.8e-83;
 Matches 274; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

 QY 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLQVLPWKTFDETIKELSRFKQVSEASV 60
 DB 1 MTEIVADKVEVVKNAIETADGALDLYNKYLQVLPWKTFDETIKELSRFKQVSEASV 60

 QY 61 LVGDIKVLMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQKIDILIRIL 120
 DB 61 LVGDIKVLMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKAQKIDILIRIL 120

 QY 121 DDGVKLNKAEQSLTSSQSFNNASGKLLALDSQLTNDFSEKSYFQSOVDRIKREAYAG 180
 DB 121 DDGVKLNKAEQSLTSSQSFNNASGKLLALDSQLTNDFSEKSYFQSOVDRIKREAYAG 180

 QY 181 AAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQNFTSLSATVQANKDIDAAK 240
 DB 181 AAAGIVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQNFTSLSATVQANKDIDAAK 240

 QY 241 LKLAETIAAIGETETETTRFYVDYDMLSLKGAACKMINTCNEYQQRHGKKTLEFV 300
 DB 241 LKLAETIAAIGETETETTRFYVDYDMLSLKGAACKMINTCNEYQQRHGKKTLEFV 300

 RESULT 5
 QYREB3 ID Q9REB3 PRELIMINARY; PRT; 305 AA.
 AC Q9REB3;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Silent hemolysin (putative pore forming hemolysin) (Hemolysin E).
 GN SHEA OR HLXE OR Z1944 OR ECS1677.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7;
 RX MEDLINE=20322319; PubMed=10865950;
 RA del Castillo F.J., Moreno F., del Castillo I.;
 RT "Characterization of the genes encoding the SheA hemolysin in
 RT Escherichia coli O157:H7 and Shigella flexneri 2a";
 RL Res. Microbiol. 151:229-230(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Pocamoussis K.,
 RA Apodaca J., Arantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayaashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shinagawa H.,
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AJ238954; CAB64962.1; -;
 DR EMBL; AE005335; AAG56033.1; -;
 DR EMBL; AP002555; BAB35100.1; -;
 DR HSSP; P77335; 1QQY.
 DR InterPro: IPR00209; Peptidase 88.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 KW Complete proteome.
 SQ SEQUENCE 305 AA; 33960 MW; 3675452D06317A45 CRC64;

Query Match 91.6%; Score 1387; DB 16; Length 305;
 Best Local Similarity 90.1%; Pred. No. 1.4e-82;
 Matches 273; Conservative 18; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MTSIPAEQTVVEVVKSAIETADGALDLYNKYLDQVVPWKTFTDIKLSRFKQYSQASV 60
 DB 3 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWKTFTDIKLSRFKQYSQASV 62
 QY 61 LVGDIKVLMDSDQKFEATQTVYEWGVVTVLLSAYILLDFEYNEKKASAKDILIRL 120
 DB 63 LVGNITKLLMDSDQKFEATQTVYEWGVVTVLLSAYILLDFEYNEKKASAKDILIRL 122
 QY 121 DGVKYLNEAQKSLTSSQFNASGKLLALDQSLNDFSEKSSYFQSQVDRIKEAYAG 180
 DB 123 DGGITKLNFAQKSLTSSQFNASGKLLALDQSLNDFSEKSSYFQSQVDRIKEAYAG 182
 QY 181 AAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVONFTSLSATVKQANKDIDAAK 240
 DB 183 AAGVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVONFTSLSATVKQANKDIDAAK 242
 QY 241 LKLAETIAAIGEIKTETTRFYVDYDMLSLKGAANKMINTCNEYQQRHGKTLFEV 300
 DB 243 LKLTETIAAIGEIKTETTRFYVDYDMLSLKGAANKMINTCNEYQQRHGKTLFEV 302
 QY 301 PDV 303
 DB 303 PEV 305

RESULT 6
 Q9X2S8 PRELIMINARY; PRT; 300 AA.
 AC Q9X2S8;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hemolysin.
 GN HLYE.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99242013; PubMed=10227474;
 RA Reingold J., Starr N., Maurer J., Lee M.D.;
 RT "Identification of a new *Escherichia coli* She haemolysin homolog in
 RT avian *E. coli*,";
 RL Vct. Microbiol. 66:125-134(1999).
 DR EMBL; AF052225; AAD28079.1; -;
 DR HSSP; P77335; 1QQY.
 SQ SEQUENCE 300 AA; 33555 MW; 6580B66C4A7B4BC CRC64;

Query Match 73.3%; Score 1110; DB 2; Length 300;
 Best Local Similarity 74.1%; Pred. No. 1.4e-64;
 Matches 215; Conservative 38; Mismatches 37; Indels 0; Gaps 0;
 QY 6 AEQTVVEVVKSAIETADGALDLYNKYLDQVVPWKTFTDIKLSRFKQYSQASV 65

DB 4 ADQTVVEVVKTAIDTADKALDLYNKYLDQVVPWKTFTDIKLSRFKQYSQASV 63
 QY 66 KVLMDSDQKFEATQTVYEWGVVTVLLSAYILLDFEYNEKKASAKDILIRL 125
 DB 64 KSLMNSQDRYFEATQTVYEWGVVTVLLSAYILLDFEYNEKKASAKDILIRL 123
 QY 126 KLNFAQKSLTSSQFNASGKLLALDQSLNDFSEKSSYFQSQVDRIKEAYAGAAAGI 185
 DB 124 KLEKAQKSLHASSQSFNSASGKLLALDQSLNDFSEKSSYFQSQVDRIKEAYAGAAAGV 183
 QY 186 VAGPFGIISYSIAAGVIEGKLIPELNNRLKTVONFTSLSATVKQANKDIDAAK 245
 DB 184 VGRPFGIISYSIAAGVIEGKLIPELNNRLKTVONFTSLSATVKQANKDIDAAK 243
 QY 246 EIAAIGEIKTETTRFYVDYDMLSLKGAANKMINTCNEYQQRHGK 295
 DB 244 EISVIGDLKTTETTRFYVDYDMLSLKGAANKMINTCNEYQQRHGK 293

RESULT 7
 Q9RMGI PRELIMINARY; PRT; 113 AA.
 AC Q9RMGI;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE HLYE.
 GN HLYE.
 OS *Shigella flexneri*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Shigella*.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12022;
 RX MEDLINE=2012345; PubMed=10660049;
 RA Wallace A.J., Stillman T.J., Atkins A., Jamieson S.J., Bullough P.A.,
 RA Green J., Artymus P.J.;
 RT "E. coli Hemolysin E (HlyE, ClyA, SheA): X-Ray Crystal Structure of
 RT the Toxin and Observation of Membrane Pores By Electron Microscopy.";
 RL Cell 100:265-276(2000).
 DR EMBL; AF200955; AAF13995.1; -;
 DR HSSP; P77335; 1QQY.
 SQ SEQUENCE 113 AA; 12879 MW; C328908D14C5C4EB CRC64;

Query Match 34.3%; Score 519; DB 2; Length 113;
 Best Local Similarity 91.0%; Pred. No. 1.1e-26;
 Matches 101; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MTSIPAEQTVVEVVKSAIETADGALDLYNKYLDQVVPWKTFTDIKLSRFKQYSQASV 60
 DB 1 MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVVPWKTFTDIKLSRFKQYSQASV 60

QY 61 LVGDIKVLMDSDQKFEATQTVYEWGVVTVLLSAYILLDFEYNEKKASA 111
 DB 61 LVGDIKVLMDSDQKFEATQTVYEWGVVTVLLSAYILLDFEYNEKKASA 111

RESULT 8
 Q9RCT3 PRELIMINARY; PRT; 115 AA.
 AC Q9RCT3;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Silent hemolysin.
 GN SHEA.
 OS *Shigella flexneri*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Shigella*.
 OX NCBI_TaxID=623;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=26;
RX MEDLINE=20322319; PubMed=10865950;
RA del Castillo F.J., Moreno F., del Castillo I.;
RT "Characterization of the genes encoding the SheA haemolysin in
RL Escherichia coli O157:H7 and Shigella flexneri 2a.";
RL Res. Microbiol. 151:229-230(2000).
DR EMBL; AJ238955; CAB65415.1; -.
DR HSP2; P77335; 10QY.
SQ SEQUENCE 115 AA; 13124 MW; 5E1F9C3449731FF8 CRC64;

Query Match      34.3%; Score 519; DB 2; Length 115;
Best Local Similarity 91.0%; Pred. No. 1.1e-26;
Matches 101; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTSIFACQTVVVKSAIETADGALDLYNKYLDQVVPKTFDETIKELSRFKQESQASV 60
DB 3 MTEIVADQTVVVKNAIETADGALDLYNKYLDQVVPKTFDETIKELSRFKQESQASV 62

QY 61 LVGDIKVLMSQDKYFEATQTVYWGCVVTVQLLSAYILLFDEYNEKKASA 111
DB 63 LVGDIKVLMSQDKYFEATQTVYWGCVATQLLAAYILLFDEYNEKKASA 113

RESULT 9
ID Q8FI27 PRELIMINARY; PRT; 93 AA.
AC Q8FI27
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Putative conserved protein.
GN C1630.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]_TaxID=217992;

SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=2238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN80095.1; -.
KW Complete proteome.
SQ SEQUENCE 93 AA; 10590 MW; CE1625028DC36DE6 CRC64;

Query Match      26.1%; Score 395; DB 16; Length 93;
Best Local Similarity 82.8%; Pred. No. 9.6e-19;
Matches 77; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 211 LNNRLKVNQFFTSLSATVQANKDIDAAKLATETAAIGETETETTRFFVYVDLDM 270
DB 1 MNKCLKSLSLFTLSNVQANKDIDAAKLATETAAIGVETETTRFFVYVDLDM 60

QY 271 LSLKGAAKMINTCNEYQHRGKKTLPFVDPV 303
DB 61 LSLKGAQAKMINTCNEYQHRGKKTLPFIEIV 93

RESULT 10
ID Q9LIW7 PRELIMINARY; PRT; 895 AA.
AC Q9LIW7
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Similar to an Arabidopsis thaliana chromosome BAC genomic

DE SEQUENCE FROM N.A.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]_TaxID=4530;
RP SEQUENCE FROM N.A.
RA Hsing Y.C., Chow T., Chen C., Wu H., Chu M., Chao Y., Liu S.;
RT "Oryza sativa PAC P0699E04 genomics sequence, complete sequence.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001111; BAA90502.1; -.
DR Gramene; Q9LIW7; -.
SQ SEQUENCE 895 AA; 100520 MW; BA4F7EFD70FFAA3 CRC64;

Query Match      8.6%; Score 130.5; DB 10; Length 895;
Best Local Similarity 22.1%; Pred. No. 2.9;
Matches 76; Conservative 66; Mismatches 123; Indels 79; Gaps 17;

QY 8 QTVEVVKSAIETADGALD-----LYNKYLDQVVPKTFDETIK-----ELSRFKQY 54
DB 274 EXVEILSEVVKLGILLDSTAESSEKRETELV--KNLESEVSVLKGLEEARIIISER 331

QY 55 SQEASVVGDIKVLMSQDKYFEATQTVYWN---CGVVTOLLAYILLFDEY-NEKKAS 110
DB 332 LAETEKLEELKSEVADAKAESEARQLFEWKKAGLLEMELEA-VTLSDKFKEGSLAS 390

QY 111 AQKDILIRLDDGVKKLNEAQKSLTSSQFNASGKLALD---SOLTNDPSEKSSYF- 166
DB 391 T-----TEELKIQSALQDRESETEVLKGTATALEIEVARLLADVNESNEQPD 438

QY 167 -----QSQVDRIRKAYAGAAAGIVAGPFGHIIISYSIAAGV-IEG-----KLPIE 210
DB 439 ASQEQEVFLQTTIDVLRNKLAAEBAEASEA-----LNNEKAANVKIEGLTEENVKLISE 492

QY 211 LN-----NRLKTVQNFFTSLSATVQANKDIDAAKLATE-----IAIGETET 257
DB 493 LNETDREKEKRAVEDLTAALS-----ESDRAKEAHERYLSKEDDHEHALAQIGDLKMAL 549

QY 258 ETTR-----FYVDYDMLSLKGAAKMINTCNEYQQR-HGKKT 296
DB 550 KSTKSEYEWLDEANVDITCLRKNVDKLEAEVKNKYRECESKET 593

RESULT 11
ID P71497 PRELIMINARY; PRT; 495 AA.
AC P71497;
DT 01-FEB-1997 (TREMELrel. 02, Created)
DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE ClpB (Fragment).
GN CLPB.
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;
OC Entomoplasmataceae.
OX NCBI_TaxID=2095;
RN [1]_TaxID=2095;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25416;
RA Falah M., Gupta R.S.;
RT "Cloning of Hsp70/DnaK gene region of Mycoplasma capricolum:
RT phylogenetic analysis of mycoplasma based on Hsp70 sequences.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U51235; AAB09427.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001270; Chaperin_c1p/b.
DR PRINTS; PR00300; CLPPTPTEASEA.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00871; CLPB_2; 1.
KW ATP-binding.
FT NON_TER 1 1
SQ SEQUENCE 495 AA; 56358 MW; A194DD51FACDF8D0 CRC64;

```

```
Query Match      8.1%; Score 122; DB 2; Length 495;
Best Local Similarity 24.6%; Pred. No. 4.7;
Matches 82; Conservative 37; Mismatches 113; Indels 102; Gaps 16;

QY 2 TSIFAEQTEVVVKSALTEADGALDLYNKVLDQVIP-----WKTDETIKELSR 49
Db 178 TGLLVKLLSSSEKERLNUEDLLKKYVKGQDAIKAVTSAMRSKGINPKDPKIGSELF 237
QY 50 F-----KQESQCE-ASVLVGBIKVLLMDSQKYFEATQV-----YEWCGVVT 91
Db 238 FQPTGVGKTEVARSLADILFNSPKKWMIRLDMSEYMEKHSVAKLIGAPPYGVYEGGRLT 297
QY 92 QLL--SAY--ILLFDEYNEKASAKQDIILRIIDGG-----VKLNEAQ 131
Db 298 EAVRRNPYSIIILFDEI-EKAHSDVENILLQILDDGRLTDSUGKTIDFNKNTIIVMTSNAS 356
QY 132 KSLTSSQSFNNASGKLLALDLSQLTNDFSEKSSYFQSQVDRIKRAYAGAAAGIVAGPFG 191
Db 357 QYLLTSDDFVQDDQKIAQ--ELNQTFRPE--FLNRIDNI--VY-----FN 396
QY 192 LIISYSIAGVIEGKLIPELNRLKTVQNFPTSLSATVKQANKDIDAAKKLATEIAAIG 251
Db 397 ALSVQTI--GEIVDKLLDELITRLQDEQNYFINFS-----EARNKIINE----- 439
QY 252 EIKTETETTRFYVDYDDL--LSLLKGAAKMINT 284
Db 440 -----GYDRLFGARLLKDLKLIET 460

RESULT 12
Q9SZK7 PRELIMINARY; PRT; 1496 AA.
AC Q9SZK7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 173 kDa protein.
GN F20D10.190 OR Ar4G38070.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Wedler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Wedler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
DR EMBL; AL035538; CAB37547.1; -.
DR EMBL; AL161592; CAB80472.1; -.
DR IncerPro; IPR001092; HLH_Basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00888; HLH_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 1496 AA; 173346 MW; 03BBB75DCD2B03FE CRC64;

Query Match      7.8%; Score 118; DB 10; Length 1496;
Best Local Similarity 19.6%; Pred. No. 36;
Matches 64; Conservative 67; Mismatches 122; Indels 74; Gaps 13;

QY 4 IPAEQTEVVVKSALTEADGALDLYNK--YLDQVIPWKT-----FDETIKELSRF-----KQ 52
Db 153 IEAEKTVKNGKNGRGGDVVVMEEKSKQVEKLVKWKKEQFKHLEAYEKLKLVKDSKK 212
QY 53 EYQESASVLVGDIKVL--LMDSQKYFEATQVYEWG--GVVTQLLSAYILLFDEYNEKKA 109
Db 213 EWEEEKSKLLDEIYSIQTKLDSVTRISEDLQKQKLMCGALQ-----EETRKHL 263
QY 110 SAKDILIRLDDGVKLINEAKSLLTSQSFNNASGKLLALDLSQLTNDFSEKSSYFQ-- 167
Db 264 EIQVSFEKAYEDAPAECDQARTQL-----DDLAKEDWEVAELRQTLMSKDAYKEM 316
QY 168 ----SQVDRIRKAYAGAAAGIVAGPFGLIISYSIAGVIEG---KLPELNRLKTVQ 220
Db 317 KYENGKLEQENRELLGSLK-----ELOEATIQSGNSALSCLKNKFNLEN 362
QY 221 FFTSLSAT-----VKQANKDIDAAKKLATEIAAIGIKTETE-----TTRFYV 264
Db 363 IHKNCANLRSKAEWSSQVEKMWVEINDYKLCQSKREALKEVELENCRSSTAKML 422
QY 265 DYDDLMLSLKGAAKMINTCNEYQQR 291
Db 423 QYEEISIMFL-----VLSRTVSEAQSR 444

RESULT 13
Q97WH8 PRELIMINARY; PRT; 587 AA.
AC Q97WH8;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE BPS2 protein homolog (bps2).
GN BPS2 OR SSO2241.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri P., Zivanovic Y., Allard G.,
RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Regan M.A., Senses C.W., Van der Oost J., Solfataricus P2.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006828; AAK42408.1; -.
KW Complete proteome.
SQ SEQUENCE 587 AA; 68436 MW; B90D5C1E19C05E86 CRC64;

Query Match      7.6%; Score 115; DB 17; Length 587;
Best Local Similarity 20.3%; Pred. No. 17;
Matches 61; Conservative 66; Mismatches 103; Indels 70; Gaps 14;

QY 10 VEVVKSALTEADGALDLYNKVLDQVIPWKTDETI--KELSRKQESQASVAVGVGIK 67
Db 42 ISLLTSSIK---AEDLLNVFADSGYVEAELDNKLXYRRIKRIINGLGEKLLIMDDRA 97
QY 68 LLMDSQKYFEATQVYEWGCVVTVQLLSA-----YLLFDEYNEKASAKQDILIRL 121
Db 98 LLL-----TYFSPENR-----LVTQILSGDGNVWFISTTSKINEIK--AKRELQKLT 145
QY 122 DGVKLINEAKSLLTSQSFNN--ASGKLLALDLSQLTNDFSEKSSYFQSQVDRIKRAY 178
```

Db 146 AEINARDELQK-----KYNIREIOAKIRAIID-----BEIDKLEK-- 181

Qy 179 AGAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFSTSLATVQKAKDIDA 238

Db 182 RESSNIVAK-----TYYTIT--LTRQNKINEILKIKVKEDELANLEFALKIEEBION 234

Qy 239 AKKLATIAAATGEIKTETETTRFRVVDYDD-----LMLSLKGAACKV-----INTCN 286

Db 235 KESKVSPIKTOLEKEMEINEKUKLTNDRSEIELEKVLVRLVEVNESDRHLDTCN 294

RESULT 14

064067 PRELIMINARY; PRT; 478 AA.

AC O64067; (TREMELrel. 07, Created)

DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)

DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)

DE Hypothetical 55.0 kDa protein.

GN YOND.

OS Bacteriophage SPBc2.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.

OX NCBI_TaxID=66797;

RN [1] SEQUENCE FROM N.A.

RA Lazarevic V., Dusterhoef A., Soldo B., Hilbert H., Mauel C., Karamata D.;

RT "The complete nucleotide sequence of the Bacillus subtilis SPbetac2

RT subphage";

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF020713; AAC13027.1; -.

KW Hypothetical protein.

SQ SEQUENCE 478 AA; 55050 MW; 6FF7495A957D4A4F CRC64;

Query Match 7.6%; Score 114.5; DB 9; Length 478;

Best Local Similarity 22.6%; Pred. No. 14;

Matches 57; Conservative 41; Mismatches 73; Indels 81; Gaps 12;

Qy 25 DLYNKY-LDQVTPWKTFDETIKELSRPKQYEQASVLVDIKVLLMSQDKYFEATQTV 83

Db 268 DYDTYFIVNVYSWS--DENSVD-KYFKFNTR-----TGDTVSIDFDSKTEVF-----MT 315

Qy 84 YEWCGVVTQLLSAYILLFDEYNKASAKQDILIRLDDGVKKLNEAQSLLTSSQSFNN 143

Db 316 RNWEEVPEPIQSOL-----NQDEQIK--DLTKQVQINKDKVGEIQENT 360

Qy 144 ASGKLALDSQ-----LTNDFSEKSYFQSQVDRIKRYKAYAGAAAGIVAG 188

Db 361 ASEKVLQNLSEVEQLPKYKEKHEKTLLEQKLSKNEFYKAKFEALNAEE----- 409

Qy 189 PFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFSTSLATVQKAKDIDA KLATEIA 248

Db 410 -----KFS-----TEVQNL-----IHASVKQ--DEEGEKAQLNTMLV 442

Qy 249 AIGEIKTETETT 260

Db 443 DLVSVPTETNTT 454

RESULT 15

031954 PRELIMINARY; PRT; 478 AA.

AC O31954;

DT 01-JAN-1998 (TREMELrel. 05, Created)

DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)

DE YOND protein.

GN YOND.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1] SEQUENCE FROM N.A.

RC STRAIN=168;

RK MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogawara N., Moser I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruchi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Daniot F., Devine K.M., Dusterhoef A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrazi E., Foulger D., Fritz C., Fujita M., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Gollightly B.J., Grandi G., Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Kiefer-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kuman M., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle B., Porwollik S., Prescott A.M., Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the gram-positive bacterium Bacillus subtilis";

RL Nature 390:249-256 (1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Kunst F., Ogawara N., Yoshikawa H., Danchin A.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z99115; CAB14031.1; -.

KW Complete proteome.

SQ SEQUENCE 478 AA; 55050 MW; 6FF7495A957D4A4F CRC64;

Query Match 7.6%; Score 114.5; DB 16; Length 478;

Best Local Similarity 22.6%; Pred. No. 14;

Matches 57; Conservative 41; Mismatches 73; Indels 81; Gaps 12;

Qy 25 DLYNKY-LDQVTPWKTFDETIKELSRPKQYEQASVLVDIKVLLMSQDKYFEATQTV 83

Db 268 DYDTYFIVNVYSWS--DENSVD-KYFKFNTR-----TGDTVSIDFDSKTEVF-----MT 315

Qy 84 YEWCGVVTQLLSAYILLFDEYNKASAKQDILIRLDDGVKKLNEAQSLLTSSQSFNN 143

Db 316 RNWEEVPEPIQSOL-----NQDEQIK--DLTKQVQINKDKVGEIQENT 360

Qy 144 ASGKLALDSQ-----LTNDFSEKSYFQSQVDRIKRYKAYAGAAAGIVAG 188

Db 361 ASEKVLQNLSEVEQLPKYKEKHEKTLLEQKLSKNEFYKAKFEALNAEE----- 409

Qy 189 PFGLIISYSIAAGVIEGKLIPELNNRLKTVQNFSTSLATVQKAKDIDA KLATEIA 248

Db 410 -----KFS-----TEVQNL-----IHASVKQ--DEEGEKAQLNTMLV 442

Qy 249 AIGEIKTETETT 260

Db 443 DLVSVPTETNTT 454

RESULT 16

09VGAS PRELIMINARY; PRT; 588 AA.

ID 09VGAS

AC 09VGAS;

DT 01-MAY-2000 (TREMELrel. 13, Created)

DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)

10-OCT-2002 (T-EMBLrel. 22, Last annotation update)
CG4066 protein.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY; PubMed=10731132;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Li X., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AF003695; AAF54775.1;
DR FlyBase; FBgn0038011; CG4066.
SQ SEQUENCE 588 AA; 63663 MW; 0CF2D6D093880B32 CRC64;
Query Match 7.6%; Score 114.5; DB 5; Length 588;
Best Local Similarity 21.1%; Pred. No. 18;
Matches 67; Conservative 38; Mismatches 97; Indels 115; Gaps 13;
QY 11 EVKSAIETADGALDLYNKYLDOV--IPWKTFTETIKELSRFKQYSQASVLVGDIKVL 68
DB 300 EILALLAENAGALDRENTQDVAISALNWKIDAE----- 333
QY 69 LMSQDKYFA-----TQVYEW-----CGVVTQLLSAYILLDFEYNEKASQ 112
DB 334 -LDNQAEYLAIEVTEBLLRNQTELLWEVLLRGVWTSFON-----LDIFANRSIEAV 387
QY 113 KDILIRILDG-----VKKLEAQAOKSLTSSQSNFNASGKLLALDSQLTNDFSEKSSV 165
DB 388 SD-LTRLQEQNKORVNLVKNVTCQILRRTKGLDD---RLNFVNQLLGLVIEPKVNS 443
QY 166 FQSDVRIKRAYAGAAAGIVAGPFGLLIISYSTAGVIEGKLIPELNRLK-----216
DB 444 LEDSFNLNKK-----SQINSILKVNPEVRNFKTSIRKLSFL 482

217 -----TVQNFFTSLSATV-----KOANK-----DIDAALKLATEATAIGEIK 254
DB 483 DNQALFENQTOENRYYSVAIVKAWPTNLKEINDLTHALSISQKRTDLAIALSGSAEYN 542
QY 255 TETETTRF--YVDYDDL 269
DB 543 TETVPTFRFISYKGIEDI 559
RESULT 17
ID Q9X360 PRELIMINARY; PRT; 652 AA.
AC Q9X360;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE PX01-90 (S-layer protein).
GN BXA0124.
OS Bacillus anthracis.
OC Plasmid virulence plasmid PX01, and Plasmid PX01.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Stearns; PLASMID=Virulence plasmid PX01;
RX MEDLINE=99445483; PubMed=10515943;
RA Okinaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hall K.K.,
RA Keim P., Koehler T.M., Lanke G., Kumano S., Mahillon J., Manter D.,
RA Martinez Y., Rieke D., Svensson R., Jackson P.J.,
RT "Sequence and organization of pX01, the large Bacillus anthracis
plasmid harboring the anthrax toxin genes."
RL J. Bacteriol. 181:6509-6515(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A2012; PLASMID=PX01;
RX MEDLINE=22061436; PubMed=12004073;
RA Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L.,
RA Holtzapple E., Busch J.D., Smith K.L., Schupp J.M., Solomon D.,
RA Keim P., Fraser C.M.;
RT "Comparative Genome Sequencing for Discovery of Novel Polymorphisms in
Bacillus anthracis."
RL Science 296:2028-2033(2002).
DR EMBL; AF065404; AAD32394.1;
DR EMBL; AF065404; AAD32394.1;
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH.
DR PROSITE; PS01072; SLH_DOMAIN; 1.
KW Plasmid.
SQ SEQUENCE 652 AA; 76210 MW; 723F5FBE03516355 CRC64;
Query Match 7.6%; Score 114.5; DB 2; Length 652;
Best Local Similarity 18.6%; Pred. No. 21;
Matches 58; Conservative 66; Mismatches 106; Indels 81; Gaps 10;
QY 25 DLYNKLDVTPWKTFTETIKELSRFKQYSQASVL-----VGDIKVLMSQDKYFE 78
DB 237 DVINQKIDE---PKNLSQRKDERMLELNQKLSLQKQSPQLQDLNKKKSQRLLE 293
QY 79 ATQ-----TVYEWCGVVTQLLSAYILLDFEYNE--KQSAQKDIILIRLDGKVK 126
DB 294 LNKXDSNRLENSIETKLNDRKAEKLLIMELIKQSEFDKIKNKEDLNKKGREDLINR 353
QY 127 LNEAQK-----SLTSSQSNFNASGKLLALDSQLTNDFSEKSSYFQ---SQV 170
DB 354 IAESEKELAKKAELNTKLVELFKVQEAALNKKSGGYLYINKLNELRELADKKAQNSNKI 413
QY 171 DRIRKEAYAGAAAGIVAGPFGLLIISYSTAGVIEGKLIPELNRLKTVQFFTSLSATVK 230
DB 414 SRLKNH-----IGEVNKLQLENIEN-----ELE 435
QY 231 QANKDIDAANKLKLAT-EIATGEIKTETETTRFVVDYDDL-----MULSLKGAAKMINT 284

436 ECNKIDNTKKQAEFDKSNKKQQLSELSVLQNLKIDELGKHKRQLEASQKXALDE 495
285 CNEYQQRHGKK 295
496 AKINKKLAEK 506

RESULT 18
Q9H6Q7 PRELIMINARY; PRT; 720 AA.
ID Q9H6Q7
AC Q9H6Q7; 2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21979 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025632; BAB15196.1;
KW Hypothetical protein.
FT NON TER 720
SQ SEQUENCE 720 AA; A86586FEAA953D0B CRC64;

Query Match 7.5%; Score 114; DB 4; Length 720;
Best Local Similarity 20.7%; Pred. No. 26;
Matches 62; Conservative 56; Mismatches 111; Indels 70; Gaps 10;

QY 26 LYNKYLDOVFPWTKFETIKELSRFKQYEQASVLVD-----IKVLLMDSQDKYFEATQ 81
DB 457 LYKEYLSEKTEWTKSTKTEKEEKLEQVQDAIKVKNLNLALQWSD-----EMKK 512
QY 82 TVVWCVVTVLLSAYILLFDEYNEKKAQAQKILIRLDDGVKLNKAEAKSLTSSQSF 141
DB 513 ILAENSKIVL-----QVNEKSLIRQYTTLVE-LERQLRKENEKQK----- 554
QY 142 NNASGKLLALDOLTNDPSEKSYFQSDVRIKAEVAGAGAGIVAGPGLIISYIAG 201
DB 555 -----ELLSWEAEV-----CEKIGCLQ-----RFKEMAFKIAA-----LQVVDNSVSLS 595
QY 202 VIE-----GLLPELNNRLKTVQNFFTLSATVKAQKIDDAKILK 242
DB 596 EELANKQYNELTAKYRDIQKQNLVORTSNLEHLECEKISLKEQVESINKLEITTEK 655
QY 243 LAT-----SIAALGEIKTETTRFVYVDYDMLSLKGAQKAMINTCNEVQQRHKK 295
DB 656 LHTEQAWQETKLGNSSMDKAKKSITNSDIVSISKKKKKKKKKKKKKKKKKKKK 714

RESULT 19
Q8CP10 PRELIMINARY; PRT; 1189 AA.
ID Q8CP10
AC Q8CP10; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chromosome segregation SMC protein.
GN S0909.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 12228;
RC Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.,

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016747; AA004506.1;
KW Complete proteome.
SQ SEQUENCE 1189 AA; 137543 MW; 0F6F95636BE31BC8 CRC64;

Query Match 7.5%; Score 114; DB 16; Length 1189;
Best Local Similarity 18.6%; Pred. No. 49;
Matches 62; Conservative 59; Mismatches 130; Indels 82; Gaps 9;

QY 7 EQTVVEVK-----SAIETADGALDLYNKYLDQVFPWKTFETIKELSRFKQYEQASV 60
DB 163 BESAGLVKRYKKRKAESIQKLDHTEDNLNRVEDILYDLGRVPELKEBAATAKEKQLSKE 222
QY 61 LVGDKVLLMDSQDKYFEATQVTVEMC-----GVVTQLLSAVILLFDEYNE 106
DB 223 MEQSDVITVSDIDHYTEDNORLDERLNHLKSOAEKEGQQAQINQLLQKY----- 273
QY 107 KKAQAQKIDILIRLDDGVKLNKAEAKSLTSSQSFNNASGKLLALDLSQLTND-----FS 160
DB 274 -KGKQQN-----DYDIEKLN--YELVKATENEQSLNVLNLEERKKNQSETNARYE 323
QY 161 EKSSYFQSDVRIKAEVAGAGAGIVAGPGLIISYIAGAGIVGKLIPELNNEL-----K 216
DB 324 EELNLSQIDSIKNEKAQNE-----KULAEKLNKQKQKLNK 359
QY 217 TVQNFFTLSATVKAQKIDDAKILKATEIAAIGEIKTETTRFVYVDYDMLSLKLG 276
DB 360 EVQELLESLLIYISDEQHEKLEIKNSVYTLMSQSDVNDIRFLEHTINEAKKSLDS 419
QY 277 AAKMINTCNEYQQ-----RHGKTKLFEV 300
DB 420 RLVEAFNLQDKIQNNITQTQKEYOSSKSKMEKV 452

RESULT 20
Q3Z429 PRELIMINARY; PRT; 538 AA.
ID Q3Z429
AC Q3Z429;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Methyl-accepting chemotaxis protein.
GN NAHY.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G7.
RX MEDLINE=95255564; PubMed=10322041;
RA Grimm A.C., Harwood C.S.;
RT "Nahy, a catabolic plasmid-encoded receptor required for chemotaxis of
Pseudomonas putida to the aromatic hydrocarbon naphthalene."
RL J. Bacteriol. 181:3310-3316(1999).
DR EMBL; AF100302; AAD13223.1; -;
DR HSP; P02942; 10U7
DR InterPro; IPR004089; Chmtaxis_trans.
DR InterPro; IPR003660; HAMP.
DR InterPro; IPR004090; Me_chemotaxis.
DR InterPro; IPR000727; T_SNARE.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR PRINTS; PS00260; CHEMTNSDUCR.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; NA; 1.
DR PROSITE; PS01111; CHEMOTAXIS_TRANSDUC_2; 1.
DR PROSITE; PS01192; T_SNARE; 1.
KW Plasmid.
SQ SEQUENCE 538 AA; 58716 MW; C91589BB65F45E56 CRC64;

Query Match 7.5%; Score 113.5; DB 2; Length 538;

Best Local Similarity					
Matches 84; Conservative 56; Mismatches 118; Indels 125; Gaps 20;					
Qy	9	TVEVVK-SAIETADGALDLYNKYLDQVIPWKTFDETIKELSRFKQEYSOEASVLVGDIKV	67	:	:
Dd	23	TIELMGFSALQR--GVASLNTVILDRVVP-----LRDLKTIADLYA-----	V 62	:	:
Qy	68	LLMDSQDK-----YFEATQTVIEWCGVVQTQLISAYILLFDEYNKKASAOK---	DILI 117	:	:
Dd	63	KIVDSHSHKARSGRMTYAQAQEVKDAGSQIDMLMHAY-----QKTKKIDEQRSVDALA	116	:	:
Qy	118	RILDGGVKKLNKAOKSLITSSQSFFNAGSKLLALDSQLTND-----FSEKSSVF-QSQ	169	:	:
Dd	117	KLVDAAQPPI-ERLKGIIL-----ERGDKAALDTFVENEMYPLIDLPSGLSHLTQIQ	167	:	:
Qy	170	VDIRIRKENAYAGAAA-----GIV--AGPFGLIISYST-----AACVI	203	:	:
Dd	168	VEE-SKRAYDAAVLYDSSRTMLALLLLGILICGGVFATRILRSIIHPHTTLTKDAAARVA	226	:	:
Qy	204	EKGKLIPEL---NNRLKTQVNFSTLSATVOAKNDIDAACKLKLATEFAAIGEKTTETET	259	:	:
Dd	227	LGDLSQSISTQVSGRNEVTDVQSVQAMQANILNTCIDOGSAAQLA---AAAEELQTATES	283	:	:
Qy	260	TR--FYVDYDDLMUS-----LLKGAACKMMINTCNE	287	:	:
Dd	284	TAQGIHRQNDEMOMATAVTEMSAAVDEADVADNAVRTSNASHAEMDLADGGRKQVMLTR	343	:	:
Qy	288	YQORHGK-----KITLFEPEDVAS	305	:	:
Dd	344	IDRLSGKLNETTRIVFLAEAS	366	:	:
 RESULT 21					
ID	Q8ZON2	PRELIMINARY;	PRT;	727 AA.	
AC	Q8ZON2;				
DT	01-MAR-2002	(TrEMBLrel. 20, Created)			
DT	01-MAR-2002	(TrEMBLrel. 20, Last sequence update)			
DE	01-OCT-2002	(TrEMBLrel. 22, Last annotation update)			
DE	Hypothetical protein A110859.				
CS	ALU0059.				
OS	Anabaena sp. (strain PCC 7120).				
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.				
QC	NCSI_TaxID=103690;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21595285; PubMed=11753940;				
RA	Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,				
RA	Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,				
RA	Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,				
RA	Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,				
RA	Yasuda M., Tabata S.,				
RT	"Complete genomic sequence of the filamentous nitrogen-fixing				
RT	cyanobacterium Anabaena sp. strain PCC 7120."				
RL	DNA Res. 8:205-213(2001).				
DR	EMBL; AP003581; BAB7583.1; "				
DR	InterPro; IPR003856; LPS_Wzz_MPA.				
DR	Pfam; PF02706; wzz. 1.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 727 AA; 82555 MW; B939994DIF4656D8 CRC64;				
 Query Match					
Best Local Similarity 22.2%; Score 113.5; DB 16; Length 727;					
Matches 75; Conservative 58; Mismatches 134; Indels 71; Gaps 15;					
Qy	5	FAEQTVEVKSIAETADGALDYLNKYLDQVIPWKTFDETIKELSRFKQEYSOEASVLVG	64	:	:
Dd	194	FINARLPVKKQLQAERKLEQFRKKHNLLDPQLSQSVLIKSLFK-TGEQLQLRFTQJOD	252	:	:
Qy	65	IKVLMDSQDKYFEATQTVIEWCGVVQTQL---SAYILLFDEYNKKASAOKDILIRILD	121	:	:
Dd	253	INSRYQNLEERTITASOK----ALVSMRLAQSSRYKLTISELOKTEOSLAKEOL-RYTD	306	:	:

```
RESULT 23
O45614 PRELIMINARY; PRT; 3102 AA.
AC O45614, P91824; Q9TZR4;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE T22A3.8 protein (Laminin alpha chain).
GN T22A3.8 OR LAMAI/2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Zhu X., Kuo G., Joh K., Hall D.H., Wadsworth V.K., Hutter H.,
RA Vogel B.E., Huang C., Yochem J., Hori K., Hedgecock E.M.,
RA Wadsworth W.G.;
RT "Expression, function and evolution of laminin alpha chains.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL008585; CAAL5432.3; -
DR EMBL; Z81125; CAB03385.3; -
DR EMBL; AL008585; CAB03385.3; JOINED.
DR EMBL; AF074902; AAC26793.1; -
DR HSSP; P02468; 1TLE
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00052; Laminin_B; 2.
DR Pfam; PF00053; Laminin_EGF; 17.
DR Pfam; PF00054; Laminin_G; 2.
DR Pfam; PF00055; Laminin_Nterm; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR ProDom; PD003031; Laminin_B; 2.
DR ProDom; PD002082; LamNT; 1.
DR SMART; SM00180; EGF_Lam; 17.
DR SMART; SM00281; LamE; 2.
DR SMART; SM00282; LamG; 4.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 12.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR PROSITE; PS00025; LAM_G_DOMAIN; 3.
KW Laminin EGF-like domain.
SQ SEQUENCE 3102 AA; 343788 MW; 2E9305D5C591FAB CRC64;

Query Match 7.5%; Score 113.5; DB 5; Length 3102;
Best Local Similarity 20.8%; Pred. No. 1.8e+02;
Matches 72; Conservative 57; Mismatches 96; Indels 121; Gaps 16;

OY 25 DLNKLVDVLPW-KTDETIKLSRFKQESQASVLVG-----DIKVLMD5 72
DB 1881 DLNKL-IDVLSQWANDRETIYVSKDTADERNLSLVVGRINRYKEVNEIEKLKRV 1939
OY 73 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104
DB 1881 QDK-----YFEATQTV-----YEWCGVVTQLLSAYILLFDEY 104

RESULT 24
Q63731 PRELIMINARY; PRT; 1999 AA.
AC Q63731;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Neuronal myosin heavy chain.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92235856; PubMed=1569576;
RA Sun W., Chantier P.D.;
RT "Cloning of the cDNA encoding a neuronal myosin heavy chain from
RT mammalian brain and its differential expression within the central
RT nervous system[see comments].";
RL J. Mol. Biol. 224:1185-1193(1992).
DR HSSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; myosin_N.
DR InterPro; IPR002928; myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; myosin_N; 1.
DR Pfam; PF01576; myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50056; IQ; 1.
KW Myosin.
SQ SEQUENCE 1999 AA; 231456 MW; 090C181D55727B41 CRC64;

Query Match 7.4%; Score 111.5; DB 11; Length 1999;
Best Local Similarity 19.6%; Pred. No. 1.4e+02;
Matches 60; Conservative 49; Mismatches 114; Indels 83; Gaps 11;

OY 7 BGTVEVTKSAIB-----TADGALDLYNKYLDVIP--WKTFBET---IKELSRFKQESQE 57
DB 1165 BGTVEVTKSAIB-----TADGALDLYNKYLDVIP--WKTFBET---IKELSRFKQESQE 57
OY 58 ASVLVGDIDKILMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNKKASAKQKILI 117
DB 1225 REELANEVNVILQGRD-----SEHKKKVKAQQLQELQ 1257
OY 118 RILDGVKKLNAQAQSLITSSQSFNNASGKLLAD---SQTNDPSEKSVFQSOVDRI 174
DB 1258 VKFNEGERVTVLADKVKLQVELDNVTGLLSQSDSKSKLTQFSALESQQLDTQELLQ 1317
```



```

RESULT 30
Q92852 PRELIMINARY; PRT; 828 AA.
AC Q92852;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Predicted coiled-coil structure containing protein.
GN MYP21540.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=1246655;
RA Sasaki Y., Ishikawa J., Yanashita A., Oshima K., Kenri T., Furuya K.,
  Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
  intracellular bacterial pathogen in humans."
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL; AF004170; BAC43945.1; -.
KW Complete proteome.
SQ SEQUENCE 828 AA; 97219 MW; 63702C9A42D3F250 CRC64;

Query Match 7.3%; Score 110.5; DB 16; Length 828;
Best Local Similarity 21.4%; Pred. No. 52;
Matches 79; Conservative 56; Mismatches 111; Indels 123; Gaps 16;

QY 2 TSIFAEQTVVVKSAIETADGALDLYNKYLD-QVIPKTEDE-TIKELSRPKQVSOEAS 59
Db 116 TFSFRKELKISADISAMQKIDISFNPIENKVIPLKNDQEQKELTALNDQKIS 175
QY 60 VLVDG-----IKVLLMSDQKFEATQTVYMGVVTQLLSAYILLDFYNE---KXA 109
Db 176 ELLVEKDIDQOSIKKHLQSNKDWENECNELV-NFSVQINQKLLISLEALIDE-NEKYTKEI 233
QY 110 SAQKIDILRIIDDGVKLINEAKSLTSSQSPFNASGKLLALDSQLTNDFSEKSYFSQ 169
Db 234 SEKNINIAHLB---KKIND---LLADIDSDNFFSK-----DFNRYNISDSF 276
QY 170 VDIRKEAYAGAAAGIVAGPGLIISYIAAGVIEGKLIPELNNRLKTVQNFFTSLSATV 229
Db 277 VSLISK-----IINNVSVDYDIKSFNAKI 302
QY 230 QKANDIDAACKLATEIA-----AIGEIKTE-----TETRFY 263
Db 303 KEFNCKKSLINKLTSSITRREHSSLSINITPDQAVYDVEESVEFANKLLSTONTKDF 362
QY 264 VDY-----DDLMLSL-----KGAACKMINTC---NEYQQRHG 293
Db 363 TDYANNFLAVPKIDTSLFGLYNELNSTLNDGVEFSKIRKINNLCIQISEYTEKN- 421
QY 294 KTLFEVPD 302
Db 422 -NLFEQVD 429

RESULT 31
Q92852 PRELIMINARY; PRT; 927 AA.
AC Q92852;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transmembrane protein.
GN LIN2460.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CLIP 11262 / Serovar 6a;
RX PubMed=11679669;
RA Glaser F., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
  Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
  Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
  Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
  Ettian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
  Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
  Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkak G.,
  Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
  Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
  Remmel B., Rose M., Schluster T., Simoes N., Tietze A.,
  Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596172; CAC97687.1; -.
DR ListList; LI02460; -.
DR InterPro; IPR005613; AIP3.
DR Pfam; PF03915; AIP3; 1.
KW Complete proteome.
SQ SEQUENCE 927 AA; 97404 MW; 7FB97E1C9E3DC54C CRC64;

Query Match 7.3%; Score 110.5; DB 16; Length 927;
Best Local Similarity 20.6%; Pred. No. 60;
Matches 79; Conservative 45; Mismatches 132; Indels 127; Gaps 16;

QY 1 MTSIFAEQTVVVKSAIETADG---ALDLYNKYLDQVIPW---KTFDETIKELSRFKOE 53
Db 160 VTKSYAEAFDKIK---ESGDGFAOAGDSGKIKDGLVKRSQENKTISTNLKTLADSSLT 216
QY 54 YSQEASVL-----VGDIK-----VLLMDS 72
Db 217 FDKGANTLVGLKTYTDGVTNTAAAGDKLNAGVSTLAAGVGLKGVAAADGATKLSGG 276
QY 73 QKRYEATQTVYMGVVTQLLSAYILLDFYNEKKSAAQKIDILRIIDDGVKLINEAK 132
Db 277 VSTYTSYGVDTL---SGGINQAYNGSTALSDGLNKWNGS-----VPALASGVTVLNDGQK 327
QY 133 SLTSSQSPFNASGKLLA---LDSQLTN-----DFSEKSYFSQVDIRKEAYAGAA 182
Db 328 SLATGLNLSLVDSGNLSAGLKELDGNLTNSQKIALQKQWMDLQGLDQLNONVNGEDA 387
QY 183 AGIVAGPGLIISYIAAGVIEGKLIPELNNRLKTVQ-----NFFTSLSAT 228
Db 388 A-----LAKQLQA---LQKSLTDLQNLGTLTKSNANPDAAIKAKINATNGVSAE 434
QY 229 VKQ-----ANKIDDAKUKLATEIAAIGEIKT-----ETETTFYVDYDDMLSL 273
Db 435 DKQKIIDAQADLDESQKSAQVATVKEKIQSGLSGLDLAATQQT-----EL 483
QY 274 LKGAACKMINTCNEYQQRHGKT 296
Db 484 QTGVAKISAG---YKAVHGGTT 502

RESULT 32
Q92852 PRELIMINARY; PRT; 979 AA.
AC Q92852;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NORPEG-like protein (Ankyrin).
GN 170002011IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuty R.K., Kuty G., Samuel W., Duncan T., Bridges C.C.,
  El-Sherbeeny A., Nagini C.N., Smith S.B., Wiggert B.;
RT "Molecular characterization and developmental expression of NORPEG: a

```


Db 73 LEQIDBEKKKART-----IKAMEAYIEDLNKIKKSDIEIKOMKIKVDINEEMK 122
QY 126 KL---NEAQSLLTSSQSPNASKGLALDSQTLNDFSEKSSYFQSQVDRIKEAYAGNA 182
Db 123 ALNISDDNRNYHTSKDKHKGISDTHDKSSKIIQNFSK-----ESDINNHIKTE----- 171
QY 183 AGIVAGPFGIIISYSAAGVIEG-KLIPELNRLKTVQNFSTLSAT----- 228
Db 172 -----LQENVSESKNSDINHVLKVENIYNILKINKIKKIIDKVKET 216
QY 229 --VKQANKDIDAKL---KLATEI---AAGEIKTETET-----RFYVDYDDIMLSL 273
Db 217 DEIKONKSINDELINGKIIKIKENSLSLTCQSKIESTIDNYSKICKIADLKTYI 276
QY 274 LKGAACKVINTCNEYQORHGKTLFEPVDP 303
Db 277 L--SEENNINTYLKNAENYENVLLNFHNI 304

RESULT 35

Q8GA40 PRELIMINARY; PRT; 1024 AA.
ID Q8GA40
AC Q8GA40
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hemolysin A.
GN HLYA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=536;
RX MEDLINE=2267134; PubMed=12379716;
RA Dobrindt U., Blum-Oehler G., Nagy G., Schneider G., Johann A.,
RA Gottschalk G., Hacker J.,
RT "Genetic structure and distribution of four pathogenicity islands (PAI
IS36 to PAI IV536) of uropathogenic Escherichia coli strain 536.";
RL Infect. Immun. 70:6365-6372(2002).
DR EMBL; AJ488511; CAD3759.1; -
SQ SEQUENCE 1024 AA; 110314 MW; 9771FOA61ECF6EB CRC64;

Query Match 7.2%; Score 109.5; DB 2; Length 1024;
Best Local Similarity 20.8%; Pred. No. 79;
Matches 70; Conservative 56; Mismatches 114; Indels 97; Gaps 11;

QY 2 TSIPAEQTVVVK-----SAIETADGALDLYNKYLDQVFPWKTFTDIKELSRFKQE 53
Db 155 TALSSMKIDELIKKQSGSNVSSSELAKASIELINQLVD-----TAASINNNVNSFSQ 208
QY 54 YSQEASVL-----VGDIKVLLMSDQKYFEATQVYEWCGVVTQLLSAYILL-FDEY 104
Db 209 LNLGSLVNTKHLNGVGN-KLQNLPLNDIGAGLDTV---SGILSASFILSNADAD 264
QY 105 NEKKAQAQKQILIRILD--GVKLNKAQKSLTSSQSPNASKGLALDSQTLNDFSE 161
Db 265 TGTAAAGVELTTKVLGVNGKISQYIIAQRAGGLSTS----- 303
QY 162 KSSYFQSQVDRIKEAYAGAAAGIVAGPFGIIIS---YSIAAGVIEGKLIPELNNRLKT 217
Db 304 -----AAAAGLIASVTLAISPLSFLSIADKFRANKIEYSORFKK 345
QY 218 VQNFSTLSATVKQANKDIDAAKLKLAETAAIGEIKTETETTFYFVYDDMLSLKGA 277
Db 346 LGYDGDLSLLAFHKTGAIDASLTITSTVLASVSSGISAATTSLV---GAPVSALVGA 401
QY 278 -----AKMINTCNEYQORHGK 294
Db 402 VTGIISGILEASKQAMFEHVASKMADVIAEWKKGK 438

RESULT 37

Q8FE01

ID Q8FE01

AC Q8FE01

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Hemolysin A.

GN HLYA OR C3570.

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O6:H1 / CFT073 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.:
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic *Escherichia coli*,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AB016766; AAN82018.1; -
KW Complete proteome.
SQ SEQUENCE 1024 AA; 110246 MW; 9DBEA50C23B5A49C CRC64;

Query Match 7.2%; Score 109.5; DB 16; Length 1024;
Best Local Similarity 20.8%; Pred. No. 79;
Matches 70; Conservative 56; Mismatches 114; Indels 97; Gaps 11;

QY 2 TSIFAEQTVVVK-----SAIETADGALDLYNKYLDQVVPKTFDETIKLSRFKQE 53
DB 155 TALSSMKIDELIKQKSGNSVSSSELAKASIELINQLVD-----TAASNNNNVSPSQ 208
QY 54 YSQEASVL-----VGDIVKVLMDSDQKYFEATQTVYEWGVTOLLQSAIYLL-PDEY 104
DB 209 LKLGSLVLSNTHLTGVGN-KLQNLPLNDNIGAGLDTV---SGILSAISAFILSNADAD 264
QY 105 NEKASAKOILIRILD-----GVKRLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSE 161
DB 265 TGTAAAGVELTTKVLGNVGKISQYIIAQRAAQGLSTS----- 303
QY 162 KSSYFQSQVDRIKAEVAGAAAGIVAGPFGIIIS-----YSTAGVIEGKLIPELNNLKT 217
DB 304 -----AAAAGLIASVVTALISFLSFLSTADKFRANKIEEYSQRFK 345
QY 218 VQNFPTSLSATVQANKDIDAAKLAETIAAIGEIKETETETTFYVDYDMLSLKGA 277
DB 346 LGYDGDLSLLAFHKTGAIDSLTISVLAHSVSGISAAATTSLV---GAPVSALVGA 401
QY 278 -----AKQMNTCNEYQQRHGK 294
DB 402 VTGIISGILEASQKAMFEHVASKRADVIAEWEKKGK 438

RESULT 38
Q8XNW6 PRELIMINARY; PRT; 1175 AA.

ID Q8XNW6
AC Q8XNW6;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 23, Last annotation update)
DE Probable exonuclease.
GN SBCC OR CFE0216.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of clostridium perfringens, an anaerobic
RT flesh-eater,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
DR EMBL; AP003185; BAB79922.1; -
DR InterPro; IPR003439; ABC transporter.
KW Exonuclease; Complete proteome.
SQ SEQUENCE 1175 AA; 136878 MW; 7C6D2366525019C1 CRC64;

Query Match 7.2%; Score 109.5; DB 16; Length 1175;
Best Local Similarity 19.3%; Pred. No. 95;
Matches 62; Conservative 65; Mismatches 121; Indels 73; Gaps 14;

QY 7 BOTVEVKSIAETADGALDLYNKYLDQVVPKTFDETIKLSRFKQ--EVSQEAS---VL 61
DB 235 EELLNKIKKEPFAEKVWNTOKELYDKRIEBSLSRSEEIKSPKERVEISNKADKIVF 294

QY 62 VGDIVKVL--MDSQ-KYFEATQTVYEWGVTOLLQSAIYLLPDEYNEK----- 108
DB 295 INNLEILKEINKDELKFSSELNKKLE-----LINLREENKLFEBFTKKKEKLPDLRLK 350
QY 109 -----ASAQDILIRILDGVKLLNEAQKSLTSSQSFNNASGKLLALDSQLTNDPSEK 162
DB 351 KEKLESQKEDILFQIKADGV-KLKEACKIFEDRSKCDTKLNSIEENKRLNEELKEK 409
QY 163 SSYFQSQVDRIK-----AVAGAAAGIVAGPFGIIISYIAAGVIEGKLIPELNNLKTQV 219
DB 410 EE-----RKEELFVHEEFKKNKINSGLF-ILNSYE-----SLDKQFNEIK 447
QY 220 NFFTSLSATVQANKDIDAAKLAETIAAIGEIKETETETTFYVDYDMLSLK---G 276
DB 448 SEEVELKYIKLITDEKSEKSEKDKLVKVESLSKINDLE-----SLLKETPG 494
QY 277 AAKMIN---TCNEYQQRHGK 294
DB 495 DNSSILEKQIKLGEYREKLNK 515

RESULT 39
C68724 PRELIMINARY; PRT; 1492 AA.

ID C68724
AC C68724;
DT 01-AUG-1998 (TREMELrel. 07, Created)
DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Phage lambda host specific protein J.
OS Yersinia pestis.
OG plasmid pMT1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=634;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM;
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott J., McCready P., Skowronski E., Garnes J.,
RA Kobayashi A., Brubaker R.R., Garcia E.,
RT "Structural Organization of virulence-associated plasmids of Yersinia
RT pestis,"
RL J. Bacteriol. 180:5192-5202 (1998).
CC -I- SIMILARITY: CONTAINS 2 FIBROECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AF053947; AAC13222.1; -
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 2.
KW Hydrolyase; Plasmid.
SQ SEQUENCE 1492 AA; 162678 MW; 3A08640BC15A9E96 CRC64;

Query Match 7.2%; Score 109.5; DB 2; Length 1492;
Best Local Similarity 19.7%; Pred. No. 1.3e+02;
Matches 57; Conservative 54; Mismatches 101; Indels 77; Gaps 9;

QY 2 TSIFAEQTVVVKSAIETADGALDLYNKYLDQVVPKTFDETIKLSRFKQSYQSEASVL 61
DB 808 SSELGOELLEEDA--KASQDAVDAINKQMEESL--KELDQSVADLDSKLEDTSGRLEGV 863
QY 62 VGDIVKVLMDSDQKYFEATQTVYEWGVTOLL-----SAIYLL 99
DB 864 QNDLKNEVSGTLDKYNDAQQVEDSNAALVELQETVSEGGKAIGAAGAAHAALDNASAL 923
QY 100 LFEYNEK-----KASA-QKDILIRILDGVKLLNEAQKSLTSSQSFNNASGKLLAL--- 151
DB 924 IABEREARVEGDKANAKQIEAMKSSVDDSAVAEEMKKTVAEVSABASASTNIEALAKT 983
QY 152 -----DSQLTND-----PSEKSSYFQSQVDRIKAEVAGAAAGIVAGPF 190
DB 984 NIDLALQDEDDQHKQVNNNAKIATQKTFADMSNASKVEIRAE----- 1029
QY 191 GLIISYIAAGVIEGKLIPELNNLKTQNFQTSISATVQANKDIDAA 239

```
Db 1030 ---IGEDIRASILE-----ETTARVEADKTIATHISKLEAQLNDDISAA 1070
RESULT 40
Q9ZH03 PRELIMINARY; PRT; 1545 AA.
AC Q9ZH03;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE Lambda host specificity protein J.
GN Y1049 OR YPWT1.06C.
OS Yersinia pestis.
OC Plasmid pMT1 (pMT-1).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM10+; PLASMID=pMT1 (pMT-1);
RX MEDLINE=99043898; PubMed=9826348;
RA Lindler L.E., Plano G.V., Burland V., Mayhew G.F., Blattner F.R.;
RT "Complete DNA sequence and detailed analysis of the Yersinia pestis
RT KIM5 plasmid encoding murine toxin and capsular antigen.";
RL Infect. Immun. 66:5731-5742(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis; PLASMID=pMT1 (pMT-1);
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyaron P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AF074611; AAC82709.1; -
DR EMBL; AL117211; CAB55188.1; -
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 2.
KW Hydrolase; Plasmid; Complete proteome.
FT VARIANT 1514 1514 G -> V.
SQ SEQUENCE 1545 AA; 168767 MW; 4E3F3B528D91DECD CRC64;

Query Match 7.2%; Score 109.5; DB 16; Length 1545;
Best Local Similarity 19.7%; Pred. No. 1.4e+02;
Matches 57; Conservative 54; Mismatches 101; Indels 77; Gaps 9;

Qy 2 TSIFAEQTVVWSALETADGALDLYNKYLDQVVPKTFETIKELSRFKQYVSQASVL 61
Db 849 SSELGQELLEIDA--KASQDAVDAINKQMESL--KELDQSVADLDSKLEDTSGRLEQV 904
Qy 62 VGDIKVLLMSQDKYFEATQTVVWCGVVTQLL-----SAYIL 99
Db 905 QNDLKNVSGTLDKNDALQVDSNALVELQETVSEQKATAGAVEAAHRAALDNASL 964
Qy 100 LFPEYNEK----KASA-QKDILIRLDDGVKYLNEAKSLTSSQSFNNASGKLAL---- 151
Db 965 IAEERAEVGDKANAKQIEAMKSSVDDSVAAVEEMKKTVAEYERASAEASTNIEALAKT 1024
Qy 152 -----DSQLND-----FSEKSSYFQSDVDRIRKEAYAGAAAGIVAGPF 190
Db 1025 NIDIALQDSDQHQWNNNAKIATQKTFFADDSAMASKVEELRAE----- 1070
Qy 191 GLIISYSIAGVIEGKLIPELNLRLKTVQNFFTSLSATVQAKNDIDAA 239
Db 1071 ---IGEDIRASILE-----ETTARVEADKTIATHISKLEAQLNDDISAA 1111

Query Match 7.2%; Score 109.5; DB 5; Length 2752;
Best Local Similarity 17.9%; Pred. No. 2.9e+02;
Matches 61; Conservative 73; Mismatches 120; Indels 87; Gaps 14;

Qy 5 FAEQTEV---VKSALETADGALDLY---NKYLDQVVPKTFETIKELSRFKQYVSQEA 58
Db 1148 XYEDPKBIKKIENVVKKDKKNIYKEINKLINEISEIEKDKTSLLEELKNINLSYGRS- 1206
Qy 59 SVLVGDIKVLMLSDQKYPFATQTVVWCGVVTQLLSAYILLFDEYNEKASAKQDILIR 118
Db 1207 ---LGNIFLEQIDEEKKABRT-----IKAMEAYIEDLNTIKGDEIEKDKWKIK 1253
Qy 119 I-LDDGVKYL---NEAQSKLTSQSFNNASGKLALDSQLTNDPSEKSSYFQSDVDRIR 174
Db 1254 MDINEEMKALNISNDDRRNVHTKSKHKGISDIHDKSSKIIQNFSK-----ESDINN 1308
Qy 175 KEAYAGAAAGIVAGPFGLIISYSIAGVIEG-KLIPELNLRLKTVQNFFTSLSAT---- 228
Db 1309 NE-----LQENVSESRKHNSDINHYSKVENYINILKINKIKKI 1347
Qy 229 -----VKQAKNDID-----AAKLKLAIEAIGETKTETTT-----RF 262
Db 1348 IDKYKEYTDEIKKNKNSINDELXNSGKILTKIK---ENSSLTEQSKISTIDNDNYISK 1404
Qy 263 YVDVDDLMLSLKGAAXKMNITCNEYQQRHGKKTLEFVDPV 303
Db 1405 IKDIADLTKTYIL--SEENNITYLKNAENYNNVLLNFHNI 1443

RESULT 42
Q9UYA4 PRELIMINARY; PRT; 595 AA.
AC Q9UYA4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Hypothetical protein PAB1294.
GN PAB1294.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
```


RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.,
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016748; AAO05028.1; -
KW Complete proteome.
SQ SEQUENCE 3692 AA; 409837 MW; 7B5FF5B21DE246F9 CRC64;

Query Match 7.2%; Score 108.5; DB 16; Length 3692;
Best Local Similarity 21.5%; Pred. No. 4.9e+02;
Matches 71; Conservative 49; Mismatches 128; Indels 83; Gaps 12;

QY 11 EVVKSATGADGALDLYNKYLQVDPWKTFDETIKELSR-----PKQESQ-EA 58
DB 564 EQVNDIIPSNVTLASYNKY-----NKLAKRAQTVDLTETNTPNQYSQTI 612
QY 59 SVLVGDIKVLMM-----DSODKYFEATQTVYEWCGVVTOLLSAYILLFDEYNEKASA 111
DB 613 DDLHELQTLINRVASREINDKACQEMTDAVYDSTELTTEKDT---LVQIENHKNEI 669
QY 112 QKDILIRILDGKVKLNKAKSLTSS-----QSFNAGSKLLALDSLTNDFS 160
DB 670 SNNDDELDDGVRKAGLHLESDTPHPVTKPNARQVNNRA-----DQCKTLIRN 723
QY 161 EKSSYFQSDVRIRK-BAYAGAAAGIVAGPFGLIISYSIAAGVIEGKLIPELNNRLKTVQ 219
DB 724 NHEATTEQNEAIRQVEARHSSDAIAKICE-----AETDTTVNEAR 763
QY 220 NFFTSLSAT-VKQANKDIDAOKLKLATIAAIGETKTETTRFVYDVLMLSLKGA 278
DB 764 DNGTKLIATDVNPFTKAE---RAAVTNSANSKIDINNNTQATLDERNDALVALVRK 820
QY 279 KKMINTCNEYQ-----QRHGKTKLPEVP 301
DB 821 DEAIQNINTAGNDVDVTEAQNNGTNTIQVP 851

RESULT 48

P96455
ID P96455 PRELIMINARY; PRT; 403 AA.
AC P96455;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Recombination protein.
GN PRE.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
SQ SEQUENCE FROM N.A.
RA Moon K.H., Lee B.R., Yoon S.J., Koh C.H., Jeong H.D., Lee D.S.;
RT "Nucleotide sequences of Rep and CAT proteins encoded by
RT chloramphenicol resistance plasmid pKH7";
RL Yakhak Hoeji 39:676-680(1995).
RN [2]
SQ SEQUENCE FROM N.A.
RA Moon K.H., Lee B.R., Yoon S.J., Koh C.H., Jeong H.D., Lee D.S.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U38429; AAB48104.1; -
DR InterPro; IPR001568; Mob_Pre.
DR Pfam; PF01076; Mob_Pre; 1.
KW Plasmid.
SQ SEQUENCE 403 AA; 47474 MW; 8620D928816DDA2A CRC64;

Query Match 7.1%; Score 108; DB 2; Length 403;
Best Local Similarity 23.2%; Pred. No. 30;
Matches 80; Conservative 45; Mismatches 130; Indels 90; Gaps 15;

QY 16 ALETADGALDLYNKYLQVDPWKTFDETIKELSRFKQESQ----- 57
DB 77 AVKHIDGLTSDSEFFDQTPEDTKQFFEHAKF--LEQEGKONLLYATVHMEKTFHM 134

QY 58 -----ASVLVGDIKVLLMSDQKYFE-ATQTVYEW-CGVVTQLLSAYILLF 101
DB 135 HYGWVPIITDDGRLSAKEVVGNKKV-LTEQDRFNEVYVNGQYDLERGQSRQVTKAKHEQI 193
QY 102 DEYNEK-----KASAKQDILIRILDGKVKLNKAKSLTSSQSFN-----NASG 146
DB 194 SQYKQTEYHKQYERESQKTDYIKOKND--KLMEYQKSLNTLKKPINVPYKQETEKVG 251
QY 147 KLIALDSQLTNDP---SEKSYFQSDVRIRKAEYAGAAAGIVAGPFGLIISYSIAAGVI 203
DB 252 GLFSKETQETGNVVISQEDFDFAKQIE-----AAQLVTEDEYIKS----- 293
QY 204 EGKLIPELNNRLKTVQNFPTSLSATVKQANKDIDAOKLKLATE-----IAAIGETKTET 259
DB 294 -EKALNDLKNSELENLENLEQIN-DIAKESQKQLNEGNKLIKVIKEFLKLEQ 351
QY 260 TRFVYVDLMLSLKXG-----AAKMINTCNEYQQRHGK 295
DB 352 AIGKSYITLMDVTKDNNQMDVLRGIDKQINPENYQEKHDIK 396

RESULT 49

Q8R8G8
ID Q8R8G8 PRELIMINARY; PRT; 549 AA.
AC Q8R8G8;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Arpase involved in DNA repair.
GN SSCC2 OR TTE2031.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
SQ SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.,
RA "A complete sequence of T. tengcongensis genome";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013153; AAM25208.1; -
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
KW Complete proteome.
SQ SEQUENCE 549 AA; 63441 MW; 19AE957652910EB7 CRC64;

Query Match 7.1%; Score 108; DB 16; Length 549;
Best Local Similarity 20.1%; Pred. No. 44;
Matches 69; Conservative 59; Mismatches 111; Indels 104; Gaps 14;

QY 5 FAEQTVGVK---SAJETAGDALDLYNKYLQVDPWKTFDETIKELSRFKQESQ----- 43
DB 175 FIEKQREILKDWEDGLKSEGNIELRQIINAKIEMKSFDETQLEKELMEQKQON 234
QY 44 ----IKELSRFKQESY-----QBASVLVGDIKVLLMD-----SQDKYFEATQTVYEW 86
DB 235 IEIEKELAILKEQYKNPPIEKPLQKDTSLQKTRBELLQYKQIQEIKNLKFPQYITCNK 294
QY 87 CG---VVT---QLLSAVILLFDEYNEKASAKQDILIRILDGKVK-----KLEAOK 132
DB 295 CGNKIDVTAKQQLLAKLQEIKEGKTAPALK-----AVIENEKAEKFEKVKRYK 350
QY 133 SLTSSQSFNNAGSKLLALDSLTNDFSKSSYFQSDVRIRKAEYAGAAAGIVAGPFG 192
DB 351 ALEIKIKRLDENALN-----TNANKQIQIQIQQIQQIQQIQQIQQIQQIQQIQQIQQI 388
QY 193 IISYSIAAGVI-----EGKLIPELNNRLKTVQNFPTSLSA-----TVKQ 231
DB 389 VISHNSVSKALLRQKEAKKLEKVEEDIKLAEQOQINQEVKTLIEYAKAFNAKLEATE 448

QY 232 ANKDIDAALKLATEIAAIGETETTRFRFYVDLMLSL 274
Db 449 INKLNKVSQQLWKIVOSTGEIKDD-----FKLYDGKEFNIL 486

RESULT 50

Q9RMD9 PRELIMINARY; PRT; 726 AA.
AC Q9RMD9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Protein tyrosine kinase.
GN WZC.
OS Acinetobacter lwoffii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=28090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RAG-1;
RA Nakar D., Gutnick D.L.;
RT "A protein tyrosine kinase of Acinetobacter lwoffii RAG-1 is involved
in emulsan biosynthesis.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ243431; CAB57193.1; -
DR InterPro; IPR003856; LPS_Wzz_MPA.
DR Pfam; PF02706; wzz; 1.
KW Kinase.
SQ SEQUENCE 726 AA; 81609 MW; 845BDEDD73E09C0E CRC64;
Query Match 7.1%; Score 108; DB 2; Length 726;
Best Local Similarity 21.7%; Pred. No. 63;
Matches 67; Conservative 49; Mismatches 91; Indels 102; Gaps 13;
QY 8 QTVEVVKSALETADGALDLYNKYLDQVTPKTT--FDETIKELSRFKQEQYSQASV----- 60
Db 173 KTDEIVLTA-----PLKLNQLKQNGFTWKVAIFTNDFDANYFVQKNSLPAAVQAILS 226
QY 61 -----LVGDIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASA 111
Db 227 NYSVAERGLTG-ILGNYGQDKEH-----ITKVLNAILAAYSQONVERRSA 273
QY 112 QKDILIRILDGV-----KKLNEAQKSLTSSQSFNNASGKLLALDSQLTNDSEKSSVFP 167
Db 274 ETAQTLKFLDEQLPDLKQLDDAERVFNKFRQYN-----IVDVTKSELYL 320
QY 168 SQVDRIR-----KEAVAGAAAGIVAGPGLIISYSIAAGVTEGKLIPELNNRLKTVQNF 221
Db 321 TQSITLETKKIELQQQQADLAAYTA-----EHPAIREINAQITALDKQ 364
QY 222 FTSLSATVKQANKDIDAALKLATEIAAIGETETTRFRFYVDLMLSLKGAOKM 281
Db 365 ITQLNSTLKQL-FDIQRQYLQLFREV-----EVKIQ-----LYTALL----- 400
QY 282 INTCNEYQQ 290
Db 401 ----NSYQQ 405

Search completed: November 28, 2003, 13:49:19
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search, using sw model

Run on:

November 28, 2003, 13:48:16 ; Search time 21 Seconds
(without alignments)
614.515 Million cell updates/sec

Title: US-09-993-292A-2

Perfect score: 1515

Sequence: 1 WPSIAEQIVVVKSAIETA.....NEYQGHKKTLFVDPVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/aaa/5A-COMB.pep:*

2: /cgn2_6/prodata/1/aaa/5B-COMB.pep:*

3: /cgn2_6/prodata/1/aaa/6A-COMB.pep:*

4: /cgn2_6/prodata/1/aaa/6B-COMB.pep:*

5: /cgn2_6/prodata/1/aaa/PCTUS-COMB.pep:*

6: /cgn2_6/prodata/1/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1396	92.1	309	1	US-08-557-115-3
2	1396	92.1	309	5	PCT-US94-05869-3
3	114	7.5	515	4	US-09-107-532A-5317
4	111	7.3	1231	4	US-09-107-532A-3150
5	108.5	7.2	1066	3	US-09-541-782-8
6	108.5	7.2	1066	4	US-09-723-820-8
7	108.5	7.2	3696	4	US-09-134-001C-5080
8	107	7.1	1211	4	US-09-134-001C-4820
9	106.5	7.0	1312	2	US-08-592-126-148
10	106.5	7.0	1312	2	US-08-687-080-51
11	106.5	7.0	1312	4	US-09-168-595-148
12	106	7.0	1315	4	US-09-328-352-4284
13	104.5	6.9	1396	4	US-09-107-532A-3944
14	103	6.8	606	4	US-08-477-831C-2
15	103	6.8	631	4	US-08-477-831C-11
16	101	6.7	477	1	US-08-402-217A-3
17	101	6.7	477	1	US-08-700-178-3
18	101	6.7	477	3	US-08-995-654-3
19	101	6.7	1329	4	US-09-310-187A-1
20	100.5	6.6	1454	4	US-09-328-352-5793
21	100	6.6	10182	4	US-09-134-001C-3159
22	98.5	6.5	1010	4	US-09-134-001C-5178
23	98	6.5	733	4	US-09-328-352-5599
24	98	6.5	1886	4	US-08-938-105-3
25	97.5	6.4	659	4	US-09-252-991A-31794
26	97.5	6.4	759	4	US-09-199-637A-170
27	97.5	6.4	2285	4	US-09-308-375-2
28	97.5	6.4	2482	1	US-08-328-254-6
29	97.5	6.4	3248	1	US-08-353-700-1
30	97.5	6.4	3248	5	PCT-US95-16216-1
31	97	6.4	961	4	US-09-914-259-66
32	96.5	6.4	457	4	US-08-924-629C-6
33	96.5	6.4	2101	1	US-08-466-390-4
34	96.5	6.4	2101	1	US-08-470-950-4
35	96.5	6.4	2101	1	US-08-467-781-4
36	96.5	6.4	2101	1	US-08-195-487-4
37	96.5	6.4	2101	2	US-08-483-924-4
38	96.5	6.4	2101	3	US-09-452-294-1
39	96.5	6.4	2101	5	PCT-US93-06160-4
40	96	6.3	975	4	US-09-914-259-19
41	95.5	6.3	1024	4	US-09-562-737-46
42	95	6.3	849	4	US-09-157-257-4
43	95	6.3	1111	1	US-08-317-450B-15
44	95	6.3	1111	3	US-08-800-593-15
45	95	6.3	1193	1	US-08-317-450B-13
46	95	6.3	1193	3	US-08-800-593-13
47	95	6.3	8991	4	US-08-714-741-32
48	94.5	6.2	779	4	US-09-934-899-8
49	93.5	6.2	442	3	US-09-081-689-2
50	93.5	6.2	442	4	US-09-305-984-14
51	93.5	6.2	442	4	US-09-073-541A-14
52	93.5	6.2	928	3	US-09-914-259-23
53	93.5	6.2	1093	3	US-09-315-793-52
54	93.5	6.2	1111	4	US-09-914-259-28
55	93	6.1	709	4	US-09-328-352-5172
56	93	6.1	789	4	US-09-107-532A-6734
57	92.5	6.1	281	4	US-09-914-259-63
58	92.5	6.1	284	4	US-09-914-259-48
59	92.5	6.1	284	4	US-09-914-259-50
60	92.5	6.1	321	4	US-09-171-517B-12
61	92	6.1	600	6	5240708-1
62	92	6.1	935	4	US-09-914-259-25
63	92	6.1	962	4	US-09-134-001C-4497
64	91.5	6.0	284	4	US-09-914-259-55
65	91.5	6.0	746	3	US-08-434-000A-4
66	91.5	6.0	746	4	US-09-312-157-4
67	91	6.0	314	4	US-09-171-517B-9
68	91	6.0	314	4	US-09-171-517B-10
69	91	6.0	3878	4	US-09-914-259-11
70	90.5	6.0	230	1	US-08-118-469A-3
71	90.5	6.0	230	1	US-08-909-119-3
72	90.5	6.0	284	4	US-09-914-259-59
73	90.5	6.0	445	1	US-08-700-359-4
74	90.5	6.0	719	4	US-09-107-532A-5992
75	90.5	6.0	927	4	US-09-107-532A-4335
76	90	5.9	904	3	US-09-198-484-2
77	90	5.9	1713	3	US-08-600-982-24
78	90	5.9	1713	5	PCT-US94-10261A-24
79	89.5	5.9	284	4	US-09-167-206-10
80	89.5	5.9	284	4	US-09-914-259-40
81	89.5	5.9	528	4	US-09-071-035-44
82	89.5	5.9	528	4	US-09-071-035-44
83	89.5	5.9	967	4	US-09-071-035-42
84	88.5	5.8	438	5	PCT-US92-09325-4
85	88.5	5.8	445	5	PCT-US92-09325-2
86	88.5	5.8	458	4	US-09-634-238-232
87	88.5	5.8	625	4	US-09-692-064-2
88	88.5	5.8	1231	4	US-08-714-741-41
89	88.5	5.8	1786	3	US-08-973-462-8
90	88	5.8	284	4	US-09-914-259-46
91	88	5.8	645	2	US-08-592-126-144
92	88	5.8	645	2	US-08-687-080-47
93	88	5.8	645	4	US-09-168-595-144
94	88	5.8	885	2	US-08-533-306A-4
95	88	5.8	885	2	US-08-742-923A-4
96	87.5	5.8	420	4	US-09-252-991A-30038
97	87.5	5.8	3075	2	US-08-460-309-5
98	87.5	5.8	3075	2	US-08-125-077-5
99	87	5.7	411	4	US-09-252-991A-23375
100	87	5.7	732	4	US-09-107-532A-6192

Sequence 6, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 6, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 19, Appli
Sequence 46, Appli
Sequence 4, Appli
Sequence 15, Appli
Sequence 13, Appli
Sequence 13, Appli
Sequence 32, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 14, Appli
Sequence 23, Appli
Sequence 52, Appli
Sequence 28, Appli
Sequence 5172, Ap
Sequence 6734, Ap
Sequence 63, Appli
Sequence 48, Appli
Sequence 50, Appli
Sequence 12, Appli
Patent No. 5240706
Sequence 25, Appli
Sequence 4497, Ap
Sequence 55, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 9, Appli
Sequence 10, Appli
Sequence 11, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 59, Appli
Sequence 4, Appli
Sequence 5392, Ap
Sequence 4335, Ap
Sequence 2, Appli
Sequence 24, Appli
Sequence 24, Appli
Sequence 10, Appli
Sequence 40, Appli
Sequence 44, Appli
Sequence 42, Appli
Sequence 21, Appli
Sequence 4, Appli
Sequence 2, Appli
Sequence 232, App
Sequence 2, Appli
Sequence 41, Appli
Sequence 8, Appli
Sequence 46, Appli
Sequence 144, App
Sequence 47, Appli
Sequence 14, Appli
Sequence 4, Appli
Sequence 30038, A
Sequence 5, Appli
Sequence 5, Appli
Sequence 23375, A
Sequence 6192, Ap

ALIGNMENTS

RESULT 1

US-08-557-115-3
; Sequence 3, Application US/08557115
; Patent No. 5731151
; GENERAL INFORMATION:
; APPLICANT: King, Harold C.
; APPLICANT: Sathish, Mundayoor
; APPLICANT: Shinnick, Thomas M.
; TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysin
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,115
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 03063-0171US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 818-3700
; TELEFAX: (404) 818-3799
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
PCT-US-08-557-115-3

Query Match 92.1%; Score 1396; DB 1; Length 309;
Best Local Similarity 91.4%; Pred. No. 2.4e-130;
Matches 275; Conservative 15; Mismatches 11; Indels 0; Gaps 0;
QY 1 MTSIFAQTVVWKSATETADGALDLYNKYLDQVTPKTDDETIKLSRFRKQYSQASV 60
DB 1 MTEIVADKTVEVWKNATETADGALDLYNKYLDQVTPKTDDETIKLSRFRKQYSQASV 60
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWGCVATQLLAAYILLFDEYNEKKSAAQKDILIRIL 120
DB 61 LVGDIKVLMDSDQKYFEATQTVYEWGCVATQLLAAYILLFDEYNEKKSAAQKDILIRIL 120
QY 121 DDGKVLNEAQSLLTSQSFNNASGKLALDQSLTNDPSEKSSYFQSDVDRIRKEAYAG 180
DB 121 DDGKVLNEAQSLLTSQSFNNASGKLALDQSLTNDPSEKSSYFQSDVDRIRKEAYAG 180
QY 181 AAAGVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
DB 181 AAAGVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
QY 241 LKLAETAAIGETITETTRFYVDYDMLSLKGAAKKMTNCNEYQORHGKTLFEV 300
DB 241 LKLAETAAIGETITETTRFYVDYDMLSLKGAAKKMTNCNEYQORHGKTLFEV 300

QY 301 P 301
DB 301 P 301

RESULT 2

PCT-US94-05869-3
; Sequence 3, Application PC/TUS9405869
; GENERAL INFORMATION:
; APPLICANT: King, C. H.
; APPLICANT: Sathish, Mundayoor
; APPLICANT: Shinnick, Thomas M.
; TITLE OF INVENTION: REGULATOR OF CONTACT-MEDIATED HEMOLYSIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 1200, The Candler Building
; STREET: 127 Peachtree Street, N.E.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05869
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414.611
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEtical: NO
; ANTI-SENSE: NO
PCT-US94-05869-3

Query Match 92.1%; Score 1396; DB 5; Length 309;
Best Local Similarity 91.4%; Pred. No. 2.4e-130;
Matches 275; Conservative 15; Mismatches 11; Indels 0; Gaps 0;
QY 1 MTSIFAQTVVWKSATETADGALDLYNKYLDQVTPKTDDETIKLSRFRKQYSQASV 60
DB 1 MTEIVADKTVEVWKNATETADGALDLYNKYLDQVTPKTDDETIKLSRFRKQYSQASV 60
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWGCVATQLLAAYILLFDEYNEKKSAAQKDILIRIL 120
DB 61 LVGDIKVLMDSDQKYFEATQTVYEWGCVATQLLAAYILLFDEYNEKKSAAQKDILIRIL 120
QY 121 DDGKVLNEAQSLLTSQSFNNASGKLALDQSLTNDPSEKSSYFQSDVDRIRKEAYAG 180
DB 121 DDGKVLNEAQSLLTSQSFNNASGKLALDQSLTNDPSEKSSYFQSDVDRIRKEAYAG 180
QY 181 AAAGVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
DB 181 AAAGVAGPFGLLISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVQANKDIDAAK 240
QY 241 LKLAETAAIGETITETTRFYVDYDMLSLKGAAKKMTNCNEYQORHGKTLFEV 300
DB 241 LKLAETAAIGETITETTRFYVDYDMLSLKGAAKKMTNCNEYQORHGKTLFEV 300
QY 301 P 301

Db 301 P 301

RESULT 3

US-09-107-532A-5317

Sequence 5317, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneka

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5317:

SEQUENCE CHARACTERISTICS:

LENGTH: 515 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...515

SEQUENCE DESCRIPTION: SEQ ID NO: 5317:

US-09-107-532A-5317

Query Match

Best Local Similarity 7.5%; Score 114; DB 4; Length 515;

Matches 81; Conservative 61; Mismatches 107; Indels 96; Gaps 19;

QY 30 YLDQVVPKTEDETIKELS-----RPKQYSQB-ASVLGDIKVLMDSDQKYFEATQT 82

Db 63 FVEVITP-----EALKEVDLSNVELLYGHDSIKSLASVKGATLK-LNVDDTGLHFEAELT 116

QY 83 -----VYEWG-GVVTQLSAYILLFDEYNEKASQAQKIDILRIILDDGYKKLNEAQS 133

Db 117 DTTYSNDVYENISKGVDSMFGFVLGLDSFDKK-----EDGTIERSDIKALNEI--S 169

QY 134 LLT-----SSQSENNASKLLALDSOL--TNDESEKSYFOSQVDRIK----- 175

Db 170 VVTVPAYSSVQVNNKRGYESFMNNOAKQTNNSLESTSKAQKESNNVEKTLIDNEKTEL 229

QY 176 ---EAYAGAAAGIVAGPGLIISYSLAAGVIEGKLIPELNRLKTVQNFETSLSATVKQA 232

Db 230 RGVEEYIRSQGEVRDG-----VTTVNAAAVVPPEVIGVEFDLKR--NNVLAQYATVKTV 282

QY 233 NKDIDAOKLKLAT-----ETAAIGEIKTETTFYDY----- 266

Db 283 SN--GQGYPVATNQAVLATKAEIAEIGDIDAEMFTS---VDYKVETRAGKIALSNEVV 337

QY 267 DDLMLSLKGA-----AKMINTCNEY-----QQRHGKKTILFEVFDV 303

Db 338 EDSAVNIVQEVKDLAKLVENDTNKHIMDLLTKTKTAATLDDL 382

RESULT 4

US-09-107-532A-5150

Sequence 5150, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneka

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5150:

SEQUENCE CHARACTERISTICS:

LENGTH: 1231 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...1231

SEQUENCE DESCRIPTION: SEQ ID NO: 5150:

US-09-107-532A-5150

Query Match

Best Local Similarity 7.3%; Score 111; DB 4; Length 1231;

Matches 78; Conservative 58; Mismatches 125; Indels 74; Gaps 17;

QY 6 AEQTVVVKSAETADGALDLYNK-----YLDQVVPKTEDE-----TIKELS 48

Db 269 ABEVIVSMGSAQTIEQTIDYLNANGRTGFLNHLIYRPPVETPVQKLPHVTVTIAYLD 328

QY 49 RPKQYSQBASVLGDIKVLMDSQ-----DKYFEATQTVYEWGCVVVTQLLSAYILLF 101

Db 329 RSKEPGAGGEPPLL-DVQSALYDSELRPAVIGRGYGLGSK-----DVTPOISA---VF 378

QY 102 DEYNKASQAQDIILIRLDGKVLKNEAQSLSLTSQSFN-----NASGKL 148
Db 379 DELKDPSTIRKRTFTIGIVDDVTYQSLPEKESLDELTEPOTQAKFWGSDGTUGANKSA 438
QY 149 LALDSQLTWDFSEKSSYFOSQ-----VDRIR-----KEAYAGAAAGTVA--GPGGLI 194
Db 439 IKIIGHDTKIAGVYFYDSSKSGGLTVSHLSPGDTPIRSAYLVHEDLVACHTP-AVILH 497
QY 195 SYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVKQ--ANKDI---DAAKLKLATEIAA 249
Db 498 SYDLVKGLKPGGIF--LLNTWSDEQLETHPLKLVLAENNIRFYTNAMRLAQEVL 555
QY 250 IGSIKETETTRF-----YVDYDLMLSLKGAKK 280
Db 556 GRRINTAMETAPFKLADIIPFDE-VLPLKEEALK 589

RESULT 5

US-09-541-782-8
; Sequence 8, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 10
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-541-782-8

Query Match 7.2%; Score 108.5; DB 3; Length 1066;
Best Local Similarity 21.6%; Pred. No. 0.092;
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;

QY 24 LDLYNKYLDQVIPWKTFDEIKELSRFKQYSEASVLVDGIKVLMDSDQKYFEATQTV 83
Db 159 LELYNEELCDLL---STDDTTK--IRFDDSTKGGSVIIQGLEEIPVHKKDDVYKLEKG 213
QY 84 YEMCGVVTQLLSAYILLFDEYNKASQAQDIILIRLDGK-----KLENAQKSLTSS 138
Db 214 KERRKTATTLMA-----QSSRSHTVSVIVHIRENGIEGEDMLKIGKLNVLDSLAGS 265
QY 139 QSFNNASGK-----LLALDSQLTN--DFSEKSSYFQSQVDRIKEAYAGAA 182
Db 266 ENVSAGNEKGIIRVETVNIQSLLTLGRVITALVDRAPHVPYRESKLTLLQESLGRT 325
QY 183 -----AGIVAGPFGL-----IISYSIAAGVIEGKLIPELNNRL--KTVQNFFTSLSATVK 230
Db 326 KTSIIATISPGHDIETLSTLEYAHRAKNIQNK--PEVNQKLTKTIVLKEYTE---EID 380
QY 231 QANKDIDAALKLATEIA--AIGHIKTETETTFYVDYDMLSLK 275
Db 381 KLKEDLMAARDKNGIYLAETGYEITLKESQNRNELKMLLKALK 427

RESULT 6

US-09-723-820-8
; Sequence 8, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015

; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

Query Match 7.2%; Score 108.5; DB 4; Length 1066;
Best Local Similarity 21.6%; Pred. No. 0.092;
Matches 62; Conservative 62; Mismatches 110; Indels 53; Gaps 12;

QY 24 LDLYNKYLDQVIPWKTFDEIKELSRFKQYSEASVLVDGIKVLMDSDQKYFEATQTV 83
Db 159 LELYNEELCDLL---STDDTTK--IRFDDSTKGGSVIIQGLEEIPVHKKDDVYKLEKG 213
QY 84 YEMCGVVTQLLSAYILLFDEYNKASQAQDIILIRLDGK-----KLENAQKSLTSS 138
Db 214 KERRKTATTLMA-----QSSRSHTVSVIVHIRENGIEGEDMLKIGKLNVLDSLAGS 265
QY 139 QSFNNASGK-----LLALDSQLTN--DFSEKSSYFQSQVDRIKEAYAGAA 182
Db 266 ENVSAGNEKGIIRVETVNIQSLLTLGRVITALVDRAPHVPYRESKLTLLQESLGRT 325
QY 183 -----AGIVAGPFGL-----IISYSIAAGVIEGKLIPELNNRL--KTVQNFFTSLSATVK 230
Db 326 KTSIIATISPGHDIETLSTLEYAHRAKNIQNK--PEVNQKLTKTIVLKEYTE---EID 380
QY 231 QANKDIDAALKLATEIA--AIGHIKTETETTFYVDYDMLSLK 275
Db 381 KLKEDLMAARDKNGIYLAETGYEITLKESQNRNELKMLLKALK 427

RESULT 7

US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 7.2%; Score 108.5; DB 4; Length 3696;
Best Local Similarity 21.5%; Pred. No. 0.63; Indels 83; Gaps 12;
Matches 71; Conservative 49; Mismatches 128;

QY 11 EVKSAIETADGALDLYNKYLDQVIPWKTFDEIKELSR-----FKQYISQ-EA 58
Db 568 EQVNDKIIPSNVTLASYNKY-----NKLKERAQTVLDBETNNTPFNQYSQTQI 616
QY 59 SVLVGDIKVLN-----DSQDKYFEATQVYEWCGVVTQLLSAYILLFDEYNKASAA 111
Db 617 DDLHELQTLINRVASREINDKAQENTDAVYDSTELTTEKDT---LVDQIENHAKNEI 673
QY 112 QKDIILIRLDGKVLKNEAQSLSLTS-----QSFNNASGKLALDSQLTNDFS 160

Db 674 SNNIDDELTDGVERVEKAGLHLESDTPHPVTKENARQVNNRA-----DQOKTLIRN 727
QY 161 EKSYSFQSVDRIRK-EAYAGAAAGIVAGPFLIISYSIAAGVIEGKLIPELNNRLKTVQ 219
Db 728 NHEATTEQNEAIRQVEAHSSDAIAKIGE-----AETDTTVNEAR 767
QY 220 NFFTSLSAT-VKQAKNDIDAALKLATEIAAIGELKTEETETFRFVDDYDMLSLKGA 278
Db 768 DNGTKLIADVPNPKKAE-----RAAVTNSANSKIKOINNNTQATLDERNDIAIALVNRK 824
QY 279 KMINTCNEYQ-----QRHKKTLFEPV 301
Db 825 DEAIQNTAQNDDVTAQNNGTNTIQVP 855
RESULT 8
US-09-134-001C-4820
; Sequence 4820, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4820
; LENGTH: 1211
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820

Query Match 7.1%; Score 107; DB 4; Length 1211;
Best Local Similarity 18.3%; Pred. No. 0.16;
Matches 61; Conservative 59; Mismatches 131; Indels 82; Gaps 9;
QY 7 EOTVEVK-----SAIETADGALDLYNKYLDQVTPKTFDETIKELSRFKQEYSQASV 60
Db 185 EBSAGLVYKKEKASIOKLDHTENLRVEDILYDLGRVPELKEEAAIAKEVYKLSKE 244
QY 61 LVGDIKVLMSQDKYFEATQTVYEW-----GVTVQLLSAYILLFDEYNE 106
Db 245 MEQSDVITVSDIDHYTEDNORLDERNLHLSQQAEGQQAQINLLQKY----- 295
QY 107 KKASAKQDILIRLDDGVKKLNEAKSLLTSSQSFNNASGKLLALDSOLTND-----FS 160
Db 296 -KGKEQQN-----DYDIEKLN---YELVKATENYEQUSGLNVLDERKKVQSETNARYE 345
QY 161 EKSYSFQSVDRIRKEAYAGAAAGIVAGPFLIISYSIAAGVIEGKLIPELNNRL-----K 216
Db 346 EBLDNLSEQIDSINKKAQNE-----KLLADLKNKQKQNLK 381
QY 217 TVQNFFTSLSATVKQAKNDIDAALKLATEIAAIGELKTEETETFRFVDDYDMLSLK 276
Db 382 EVOELESILYSDSEQDKLEIKNSYYTLMSQSVNNDIRFLEHTINENEAKKSLDS 441
QY 277 AAKMINTCNEYQ-----RHGKTLFEPV 300
Db 442 RLVEAFNOLKDIQQNITOTQKEYQSSKSMKV 474

RESULT 9
US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory

; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54
; US-08-592-126-148..

Query Match 7.0%; Score 106.5; DB 2; Length 1312;
Best Local Similarity 18.9%; Pred. No. 0.2;
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;
QY 10 VEYVKSATETADGALDLYNKYLDQVTPKTFDETIK-----ELSRFKQY 54
Db 502 MEVISLQNEKAD--LDRTRKLDQMEQLNHHHTTRQVEMLTQKAKDKQDEIRKIKSRH 559
QY 55 SOEASVVGDIKVLMSQDKYFEATQTVYEWGVTVQLLSAYILLFDEYNEKKASQKD 114
Db 560 SDELTSLG-----YFPNKKQLEDWLHKSCK-----EINQTR----- 591
QY 115 ILIRLDDGVKKLNEAKSLLTSSQSFNNASGKLLALDSOLTNDPSEK-----SSYFQS 168
Db 592 -----DRLAKLN---RELASSEQNKHNNELKRREBEQJSS--YEDKLFVCGSQDPES 640
QY 169 QVDRIRKE-----AYAGAAAGIVAGFF----- 190
Db 641 DLDRLEEIEKSSKQRAVLGATA--VYSQFITQLTDENQSCCPVCQVQFQTEAELOEVI 698
QY 191 -----GLI-ISYSIAAGVIEGKLIPELNNRLKT 217
Db 699 SLDQSKRLAPDKLSTESSELKKKRRDEMGLVPMRQSIID--LKEKEIPELNNKLQN 756
QY 218 VQNFFTSLSATVKQAKNDIDAALKLATEIAAIGELKTEETET-----RFYVDY 266
Db 757 V-----NRDIQRLKNDIEBQETLLGTIMPEESAKVCLTDVTIMERFQEMEL 802
QY 267 DDLMSLLKGAAG-----KMINTCNEYQQRHGKK 295
Db 803 KQVERKIAQQAARLQGLDLDRTVQVQVNOEKQKQHK 838
RESULT 10
US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427

GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 7.0%; Score 106.5; DB 2; Length 1312;
Best Local Similarity 18.9%; Pred. No. 0.2; Indels 169; Gaps 16;
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;

Qy 10 VEVVKSATETADGALDLYNKYLDQVVPWKTFTETIK-----ELSRFKQY 54
Db 502 MEVSLQNEKAD--LDRTLRLKLDQEMQLNHHHTTQTQEMTLTKDKADKDEQIRKIKSRH 559
Qy 55 SQEASVLVGDIKVLLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEVNEKASAKQD 114
Db 560 SDELTSLG-----YFNNKQLEDLHLSKSK-----EINQTR----- 591
Qy 115 ILIRILDGVKKLNEAOKSLTSSQSPFNAGSKLLALDSQLTNDPSEK-----SSYFQS 168
Db 592 -----DRLAKLN---KELASSEQNKHNINNELKREKEQLSS--YEDKLFVDCGSDQFES 640
Qy 169 QVDRIRKE-----AYAGAAAGIVAGPF-----190
Db 641 DLDRLEEIEKSSKQRAMLAGATA--VYSQFITQLTDENSCCPVQCRVQTEAELOEVI 698
Qy 191 -----GLI-YSYISIAAGVIEGKLIPELNNRLKT 217
Db 699 SDLQSKLRAPDKLKSTESLKKKERRDEMLGLVPMRQSIID--LKEKEIPELRNKLQN 756
Qy 218 VQNEFTSLSATVQKQNDIAAKLKATEIAAGTEIKTETETT-----RYVDY 266
Db 757 V-----NRDIQRLKNDIEBQETLLGTIMPEESAKVCLTDVTIMRPFQNEL 802
Qy 267 DDLMLSLIKGAAK-----KMINTCNEYQQRHGCK 295

Db 803 KQVERKIAQQAQAKLQIGIDLDRTVQVQVQKQKQHK 938

RESULT 11
US-09-168-595-148
Sequence 148, Application US/09168595
Patent No. 6555666
GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/168,595
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/592,126
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEtical: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
US-09-168-595-148

Query Match 7.0%; Score 106.5; DB 4; Length 1312;
Best Local Similarity 18.9%; Pred. No. 0.2; Indels 169; Gaps 16;
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;

Qy 10 VEVVKSATETADGALDLYNKYLDQVVPWKTFTETIK-----ELSRFKQY 54
Db 502 MEVSLQNEKAD--LDRTLRLKLDQEMQLNHHHTTQTQEMTLTKDKADKDEQIRKIKSRH 559
Qy 55 SQEASVLVGDIKVLLMDSQDKYFEATQTVVWCGVVTQLLSAYILLFDEVNEKASAKQD 114
Db 560 SDELTSLG-----YFNNKQLEDLHLSKSK-----EINQTR----- 591
Qy 115 ILIRILDGVKKLNEAOKSLTSSQSPFNAGSKLLALDSQLTNDPSEK-----SSYFQS 168
Db 592 -----DRLAKLN---KELASSEQNKHNINNELKREKEQLSS--YEDKLFVDCGSDQFES 640
Qy 169 QVDRIRKE-----AYAGAAAGIVAGPF-----190
Db 641 DLDRLEEIEKSSKQRAMLAGATA--VYSQFITQLTDENSCCPVQCRVQTEAELOEVI 698
Qy 191 -----GLI-YSYISIAAGVIEGKLIPELNNRLKT 217
Db 699 SDLQSKLRAPDKLKSTESLKKKERRDEMLGLVPMRQSIID--LKEKEIPELRNKLQN 756

; STATE: CA
 ; COUNTRY: USA
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/402,217A
 ; FILING DATE: 10-MAR-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Luther, Barbara J.
 ; REGISTRATION NUMBER: 33954
 ; REFERENCE/DOCKET NUMBER: PF-0028US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-852-0195
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 477 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; IMMEDIATE SOURCE:
 ; LIBRARY: mouse
 ; CLONE: GI 53979
 ; US-08-402-217A-3

Query Match 6.7%; Score 101; DB 1; Length 477;
 Best Local Similarity 20.1%; Pred. No. 0.15;
 Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;
 QY 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLDVFPWKTFTDETIKELSRFQKQYSEASV 60
 Db 99 LNNLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
 QY 61 LVGDIKVLMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120
 Db 156 L-RDVTAAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQSIRDVTAQ-L 205
 QY 121 DGVKVLNKAQKSLTSS---QSFNNAQKLLALDSQLTN-DFSEKSSYFQSQVDRIKKE 176
 Db 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIE-DLKLENLTQEKVMAEKSVEDVQQQ 264
 QY 177 AVAGAAAGIVAGPFGLLIISYTAAGVIEGKLIPELNNR-----LKTQVQNF---TSL 225
 Db 265 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETSSFLEKITDL 307
 QY 226 SATVKQANKDI-----DAKKLKATEIA-----AIGIKTETETTRFYVD-Y 266
 Db 308 KNQLRQODEDFRQLEKKGRTAEKENVMTELTMEINKWRLLYEELYEKTKPFQOQLDAF 367
 QY 267 DDLMLSLK--GAACKWINTONE-YOORHGKTL 297
 Db 368 EAEKQALLNEHGATQEQLNKIRDSYAQLLGHQNL 401

RESULT 17
 US-08-700-178-3
 ; Sequence 3, Application US/08700178
 ; Patent No. 5783669
 ; Patent No. 5783669 5700912
 ; GENERAL INFORMATION:
 ; APPLICANT: Hawkins, Phillip R.
 ; APPLICANT: Wilde, Craig G.
 ; APPLICANT: Seilhamer, Jeffrey J.
 ; TITLE OF INVENTION: HYALURONAN RECEPTOR EXPRESSED IN HUMAN
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/700,178
 ; FILING DATE: August 20, 1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/402,217
 ; FILING DATE: March 10, 1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0028-1 DIV
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 477 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; IMMEDIATE SOURCE:
 ; LIBRARY: mouse
 ; CLONE: GI-53979
 ; US-08-700-178-3

Query Match 6.7%; Score 101; DB 1; Length 477;
 Best Local Similarity 20.1%; Pred. No. 0.15;
 Matches 67; Conservative 72; Mismatches 127; Indels 68; Gaps 18;
 QY 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLDVFPWKTFTDETIKELSRFQKQYSEASV 60
 Db 99 LNNLREKEVELEKHIAHAQAAILIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
 QY 61 LVGDIKVLMSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNEKKASAKDILIRIL 120
 Db 156 L-RDVTAAQLESQEKYNDTAQSLRD---VTAQLESEQ---EKYND-TAQSIRDVTAQ-L 205
 QY 121 DGVKVLNKAQKSLTSS---QSFNNAQKLLALDSQLTN-DFSEKSSYFQSQVDRIKKE 176
 Db 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIE-DLKLENLTQEKVMAEKSVEDVQQQ 264
 QY 177 AVAGAAAGIVAGPFGLLIISYTAAGVIEGKLIPELNNR-----LKTQVQNF---TSL 225
 Db 265 ILTAESTN-----QEYA-----RMVQDLQNRSTLKEEIKETSSFLEKITDL 307
 QY 226 SATVKQANKDI-----DAKKLKATEIA-----AIGIKTETETTRFYVD-Y 266
 Db 308 KNQLRQODEDFRQLEKKGRTAEKENVMTELTMEINKWRLLYEELYEKTKPFQOQLDAF 367
 QY 267 DDLMLSLK--GAACKWINTONE-YOORHGKTL 297
 Db 368 EAEKQALLNEHGATQEQLNKIRDSYAQLLGHQNL 401

RESULT 18
 US-08-995-654-3
 ; Sequence 3, Application US/08995654
 ; Patent No. 6025138
 ; GENERAL INFORMATION:
 ; APPLICANT: Hawkins, Phillip R.
 ; APPLICANT: Wilde, Craig

APPLICANT: Seilhamer, Jeffrey
TITLE OF INVENTION: HVALORAN RECEPTOR EXPRESSED IN HUMAN
TITLE OF INVENTION: UMEILICAL VEIN ENDOTHELIAL CELLS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/995,654
FILING DATE: December 22, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/700,178
FILING DATE: August 20, 1996
APPLICATION NUMBER: 08/402,217
FILING DATE: March 10, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PP-0028-2 DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-995-654-3

Query Match 6.7%; Score 101; DB 3; Length 477;
Best Local Similarity 20.1%; Pred. No. 0.15; Mismatches 127; Indels 68; Gaps 18;
Matches 67; Conservative 72;
QY 1 MTSIFAQTVVVKSAIETADGALDLYNKYLDQVTPWKTFTETIKELSRFKQYSQEA5V 60
DB 99 LDNLLREKEVELEKXIAHAQAIIAQEKYNDTA---QSLRDVTAQLESVQEKYNDTAQS 155
QY 61 LVGDIKVLMDSDQKYFEATQTVYEWGCVTQLLSAYILLFDEVNEKKAQAQDIILIRIL 120
DB 156 L-RDVTAQLESQEKYNDTAQSLRD---VTAQLESEQ-----EKYND-TAQLRDVTAQ-L 205
QY 121 DGVKVKNEAKSKLLTS---QSFNNAAGKLLADSLTN-DFSEKSSYFQSQVDRIKRE 176
DB 206 ESVQEKYNDTAQSLRDVSAQLESYKSTLKEIE-DLKLNLTLQEKVMAEKSVEDVQQQ 264
QY 177 AVAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNR-----LKTQVQNF-----TSL 225
DB 265 ILTAESTV-----QEYA-----RVQDLQNRSLKEEIEKITSSLEKITDL 307
QY 226 SATVQAQNKDI-----DAAKLKLATEIA-----AIGEIKTETTRFFVYD-Y 266
DB 308 KNQLRQDEDPFKQLEEKGRKTAENVMVTELTWEINKWRLLYBEELYEKTTPQQLDAP 367
QY 267 DDLMLSLK--GAAKKMTNCNE-YQQEHGKTL 297
DB 368 EAEQALLNEHGATQEQQLNKIRDSYAQLLGHQNL 401

RESULT 19
US-09-310-187A-1

Sequence 1, Application US/09310187A
Patent No. 6388751
GENERAL INFORMATION:
APPLICANT: Benichou, Gilles
APPLICANT: Fedoseyeva, Eugenia
TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
TITLE OF INVENTION: Graft Rejection
FILE REFERENCE: UCSF-090
CURRENT APPLICATION NUMBER: US/09/310,187A
CURRENT FILING DATE: 1999-05-12
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1939
TYPE: PRT
ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match 6.7%; Score 101; DB 4; Length 1939;
Best Local Similarity 22.5%; Pred. No. 1.3; Mismatches 138; Indels 86; Gaps 13;
Matches 75; Conservative 44;
QY 5 FAQTVVVKSAIETADGALDLYNKYLDQVTPWKTFTETIKELSRFKQYSQEA5VLVD 64
DB 1173 FQMRDRLEEAATLQHEATAALRKHADSV---AELGEQIDNLRQVKLEKES-----E 1225
QY 65 IKVLLMSQDKYFEATQTVYEWGCVTQLLSAYILLFDEVNEKKAQAQDIILIRILDDGV 124
DB 1226 FKLLEDD-----VTSNMEQII-----KAKANLEKVSRTLEDQA 1258
QY 125 K-----KLENAQKSL---LTSSQSFNNAAGSKLL-----ALDSQITNDFSEKSSYFQSQV 170
DB 1259 NEYRVKLEEARQSRSLNDFTTQRAKIQTENGEARQLEKEKALISQLTRG---KLSYQOME 1315
QY 171 DRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPE-----LNNRL 215
DB 1316 DLKQQLBEEGKAKNALA-----HALQSAHDCDLLEQYEBETEAKELQVLSKAN 1367
QY 216 KTVQNFFTSLSATVQAQNKDIADAKLKLATE-----IAAIGIKTETET--RFVYDY 266
DB 1368 SEVAQWRKYETDAIQRTTEEELEAKKLAQRLQDAEAVEAVNAKSSLEKTKRLQNEI 1427
QY 267 DDLMLSLKGAAGKMTNCNEVQQRHGKTLFE 299
DB 1428 EDLMDTVERSNAA---NALDKQNFPKILAE 1457

RESULT 20

US-09-328-352-5793
Sequence 5793, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Bretton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5793
LENGTH: 1454
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5793

Query Match 6.6%; Score 100.5; DB 4; Length 1454;
Best Local Similarity 19.6%; Pred. No. 0.93; Mismatches 133; Indels 139; Gaps 18;
Matches 80; Conservative 57;
QY 11 EYVKSIAET-----ADGALDLYNKYLDQVTPWKTFTETIKELSRFKQYSQEA5V 50
DB 278 QVLAGSVQTLASNLDIADGALVGVIGYITRAILMKSS-AAIKEGMASLASRQASVLNA 335

```

QY 51 KQBYQERASVL-----VGDIKVILMDSQKYFEAT-----QTVE 85
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 QAEYAEATAALNAAKAHLANRATNAETOAK-FCGATAATRYAQAAVTAATNAQTAQ 394

QY 86 -----WGVVVTQLLSAYILLPDEYNKKAQKDI-----LI 117
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 IKLNTATSIAGRLAGAFGLIGWAGVATLGWGLAAAYFPNNKAEAKOKLAQAKVA 454

QY 118 RILDGVKVL--NEAQKSLTSSOSFNNAKGLIALDSQLTNDF----SEKSYFOSQVDR 172
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 EXADEBELKKTJNDKAKAVNDLTAF-NAQNKALEKSSRAVGSALIDIEYARGNREVK 513

QY 173 IRKEYAGAAGIVAGPFGELIYSISIAAGVIEGKLIPELNNRLKTVQNETFSLSATVKQ- 231
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 514 IQGEARTGT-----ISYTEAI-----ERLNKIKLPDLYENLKQAAQY 552

QY 232 ----ANKDIDAAKKL-----ATEIAAIGETETE-TTRFVVDY- 266
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 553 DNNASKASLSAEKLLRREVYKLGNEFAQNAATQHQAQDALGNTATEAEKATKALQDYQ 612

QY 267 -----DDLMLS--LLKGAAKQWINTCNEYQQRHGKTLFVFPDVAS 305

Db 613 AKQKDSVIDSIYKSGWLQKGYTVACANAILELOKAKGMSAILSKDRIDS 661

```

```

RESULT 21
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYNN Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

```

```

Db      3850 VQAKQQLQ-----QIVNEVDPTQ 3867

RESULT 22
US-09-134-001C-5178
; Sequence 5178, Application US/091344001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AMINOC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGN
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

```

```

RESULT 23
US-09-328-352-5599
; Sequence 5599, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Brston et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5599
; LENGTH: 733
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5599

```

Best Local Similarity 22.0%; Pred. No. 0.57;
Matches 65; Conservative 43; Mismatches 86; Indels 102; Gaps 14;
QY 12 VKSATETAGADLNNKYLQVDPKTFDETIKELS-----RFQKYS-QEASVIVGD 64
Db 201 VKKVAIFTKD---OLDSKY-----NIKLSIPAADVNISSNYSVAERGLTG- 244
QY 65 IKVLLMDSQDKYFEATQTVYVWCGVVTQLLSAVILLFDEYNEKKASAKQDILIRILDDGV 124
Db 245 ILGLNYQGYDKEH-----ITQVLNAILVYGAQNVERRSASQTLKFLDEQL 292
QY 125 -----KLNKSAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSYFQSQVDRI- 174
Db 293 PDLKQDDDAEROFNFRQOYN-----TVDVTKESELYLTOSITLTKAEL 339
QY 175 KEAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIPELNNRLKTVQNFSTLSATVKQANK 234
Db 340 EQQAEWAAYTA-----EHPAMREINGQITAINQIGELNSTLKQL-P 382
QY 235 DIDAAKKLATEIAAIGEIKETETTRFYVDYDMLSLKGAAGKMMINTCNEYQQ 290
Db 383 DVQRQYLQLYREV-----EVRTQ-----LYTALL-----NSYQQ 411

RESULT 24

US-08-938-105-3

; Sequence 3, Application US/08938105

; Patent No. 6353151

; GENERAL INFORMATION:

; APPLICANT: Weinwand, Leslie A.

; APPLICANT: Vikstrom, Karen L.

; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross P.C.

; STREET: 1700 Lincoln St., Suite 3500

; CITY: Denver

; STATE: CO

; COUNTRY: U.S.A.

; ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/938,105

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Crook, Wanneil M.

; REGISTRATION NUMBER: 31,071

; REFERENCE/DOCKET NUMBER: 3595-4

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 863-9700

; TELEFAX: (303) 863-0223

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1886 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-938-105-3

Query Match

Best Local Similarity 6.5%; Score 98; DB 4; Length 1886;

Matches 71; Conservative 44; Mismatches 131; Indels 88; Gaps 12;

QY 5 FASQTVWVKASATADGALDLYNKYLQVDPKTFDETIKELSREKQYSQEASVIVGD 64
Db 1120 FQWRDLBATQHEATAALKKGRADSV---AELGEQDNLQRVYQKLEKES-----E 1172
QY 65 IKVLLMDSQDKYFEATQTVYVWCGVVTQLLSAVILLFDEYNEKKASAKQDILIRILDDGV 124

Db 1173 FKLEDD-----VTSHMEQII-----KAKANLEKVSRTLEDQA 1205
QY 125 K-----KLNKSAQKSLTSSQSFNNASGKLLALDSQLTNDPSEKSYFQSQ 169
Db 1206 NEYRVKLEEAQRSL-----NDFTTQRAKLQTEGELARQLEEKALIQWLRGKLSYQQM 1261
QY 170 VDRIRKEAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIPE-----LNNR 214
Db 1262 EDLRQQLKEEGKAKVALA-----HALQSAHDCCLLEQYEEEAELQRVUSKA 1313
QY 215 LKTQVNTFTLSATVKQANKDIDAALKLATE-----IAAIGEIKETETTT--RFYVD 265
Db 1314 NSEVAQWRTKYETDAIQTELEEAQKQLAQRLQDAEAEVAVNAKCSLEKTKHRLQNE 1373
QY 266 YDDMLSLKGAAGKMMINTCNEYQORHGKKTLP 299
Db 1374 IEDLMVDVERSNA--AALDKKQNFDKILAE 1404

RESULT 25

US-09-252-991A-31794

; Sequence 31794, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 31794

; LENGTH: 669

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; US-09-252-991A-31794

Query Match 6.4%; Score 97.5; DB 4; Length 669;
Best Local Similarity 21.8%; Pred. No. 0.56;
Matches 57; Conservative 48; Mismatches 94; Indels 63; Gaps 10;

QY 33 QVTPKTFDETIKELSRFKQY-----SQEASVL-----VGDIKVLLMDSQDK 75
Db 112 EVFPGKQVDDIDAEIALFKQALEGVRADMRALSKLASQLRKBERALFDVYLMMLDDASI 171
QY 76 YFEATQTVY--EWC-GVVTQLLSAYI---LLFDEYNEKKASAKQDILIRILDDGVKLN 128
Db 172 GNEVKRIIRTQWAGALRQVMEHVQRFELMDDAYLREASDVKDIGRALL-----AYLQ 227
QY 129 EAQKSLT-----SQSFNNA-----SGKLLALDSQLTNDPSEKSYFQSQVDRIK 175
Db 228 EERKQNLTYPEQTLIVSEELSPAMLGEVPEGRVLGVSVLGS-----NSHVALIAR 279
QY 176 EAYAGAAAGIVAGPFGLLIISYIAAGVIEGKLIPELNNRLKTVQNFSTLSATVKQANKD 235
Db 280 AMGIPTVMGAVDLPYSKVVDIGDIIVDGYHGEV-----YTNFSABELVROYSD 325
QY 236 IDAAKLKLAETIAAIGEIKETET 257
Db 326 VVAERELSKGLAALRELPCET 347

RESULT 26

US-09-139-637A-170

; Sequence 170, Application US/09199637A

; Patent No. 6355411

; GENERAL INFORMATION:

; APPLICANT: Ausubel, Frederick

```
Query Match      6.4%; Score 97.5; DB 1; Length 2482;
Best Local Similarity 20.5%; Pred. No. 4.2;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;
```

```
Query Match      6.4%; Score 97.5; DB 4; Length 2285;
Best Local Similarity 19.9%; Pred.No 3.7;
Matches 65; Conservative 58; Mismatches 106; Indels 97; Gaps 16;

QY          9 TVEVVKSAIEADAGL-----DLNKKYLDOVIPAKTFDFDTIKELSRFKQEVSQE 57
             |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```


QY 55 SQASVULGDIKVLMD---SQKYFEAT-----QTVYHCGVVTQLLSAY 97
DB 2008 SGEKNRAGLQQLLEIKSKDQLKELTLENSELKSLDCMHKQDQVEKGVREEIABY 2067
QY 98 ILLFDEYNEKASA-----QKDILIRILDDGVKKLNEAQS-----LITSSQSFNN 143
DB 2068 QLRHBE-AEKHQALLDNTKQVEVEIQTYREKLTSEKCLSSQKLEIDLKSSKEELNN 2126
QY 144 ASKLLALDSQLTNDFSEKSSYFQSQVDRIKREYAGAAAGIVAGPFGILLIISYSIAAGVI 203
DB 2127 S-----LKATTTQILEELKTKMDNLKYVNLKENER-----AQGKMKLLI----- 2167
QY 204 EGKLIPELNNRLKTVQNFSSL-SATVQKANKIDDAKL-KLATEIAAIGE-IKTETETT 260
DB 2168 --KSCQLEBEKEILQKELSQLQAQEKQKTGTVMDTKVDLTTTELKELTELEKTKEA 2225
QY 261 RFYVD-YDDLMLSLK-GAAKQKMT-----CNEYQORHGK 294
DB 2226 DEYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQSQSKQDSR 2266

RESULT 29
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
US-08-353-700-1

Query Match 6.4%; Score 97.5; DB 1; Length 3248;
Best Local Similarity 20.5%; Pred. No. 6.3;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

QY 7 EQTVVVKSAITETADGALDLYN-----KYLDQVIPW-----KTFDETIKELSRFKQEV 54

DB 2676 QDTLEVLQSSYKLNLELELTMDKKSQSFVEKVNMTAKETELOREMHEMAQKTAELQEEL 2735
QY 55 SQEASVULGDIKVLMD---SQKYFEAT-----QTVYHCGVVTQLLSAY 97
DB 2736 SGEKNRAGLQQLLEIKSKDQLKELTLENSELKSLDCMHKQDQVEKGVREEIABY 2795
QY 98 ILLFDEYNEKASA-----QKDILIRILDDGVKKLNEAQS-----LITSSQSFNN 143
DB 2796 QLRHBE-AEKHQALLDNTKQVEVEIQTYREKLTSEKCLSSQKLEIDLKSSKEELNN 2854
QY 144 ASKLLALDSQLTNDFSEKSSYFQSQVDRIKREYAGAAAGIVAGPFGILLIISYSIAAGVI 203
DB 2855 S-----LKATTTQILEELKTKMDNLKYVNLKENER-----AQGKMKLLI----- 2895
QY 204 EGKLIPELNNRLKTVQNFSSL-SATVQKANKIDDAKL-KLATEIAAIGE-IKTETETT 260
DB 2896 --KSCQLEBEKEILQKELSQLQAQEKQKTGTVMDTKVDLTTTELKELTELEKTKEA 2953
QY 261 RFYVD-YDDLMLSLK-GAAKQKMT-----CNEYQORHGK 294
DB 2954 DEYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQSQSKQDSR 2994

RESULT 30

PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 6.4%; Score 97.5; DB 5; Length 3248;
Best Local Similarity 20.5%; Pred. No. 6.3;
Matches 70; Conservative 70; Mismatches 126; Indels 75; Gaps 16;

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,390
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

Query Match 6.4%; Score 96.5; DB 1; Length 2101;
Best Local Similarity 22.6%; Pred. No. 4.1;
Matches 65; Conservative 50; Mismatches 107; Indels 65; Gaps 14;
QY 42 ETIKELSRFKQVYSQASVLVDIKVLLMDSQDKYFEATQVYEWCGVVTQLLSAYILLF 101
DB 294 ETLKQCQDLKTEKSQ-----MDRKINQISENGDLSFK-----LREFASHLQQLDALNELT 345
QY 102 DEYN-----EKASAKQDILIRLDGKVKLNEAKSLTSSQSFNNASGKLLALDSQ 154
DB 346 EEHSKATQEWLEKQAKLEKLSAALQD---KKCLEEKNEIL-----QKLSQLEEH 393
QY 155 LT-----NDFSEKSSYFQS--QVDRIRKEAYAGAAGIVAGPFGI---IISYSIAAGVIEG 205
DB 394 LSQIQDNPPEKGEVLGDVLETLKQE-----AATLAANTQLOARVEMLETERGQGEA 448
QY 206 KLIFE-----LNRLKTVQNFSTSATVKQANKDIDAUKLATEIAA-IGEIK 254
DB 449 KLLAERGHFEKQKLSLITDQSSISNLS---QAKEELEQASQAHGARLTAQVASLT 504
QY 255 TETETTRFYDDYDMLSLKGAAGK---XMINTCNEYQO-----RH 292
DB 505 SELTILNATIOQQQBELAGLQKQAKEQAQLAQTLOQQEQASQGLRH 551

RESULT 34
US-08-470-950-4
Sequence 4, Application US/08470950
Patent No. 5698439
GENERAL INFORMATION:
APPLICANT: TOUKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HUWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,950
FILING DATE: 06-JUN-1995

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-950-4
Query Match 6.4%; Score 96.5; DB 1; Length 2101;
Best Local Similarity 22.6%; Pred. No. 4.1;
Matches 65; Conservative 50; Mismatches 107; Indels 65; Gaps 14;
QY 42 ETIKELSRFKQVYSQASVLVDIKVLLMDSQDKYFEATQVYEWCGVVTQLLSAYILLF 101
DB 294 ETLKQCQDLKTEKSQ-----MDRKINQISENGDLSFK-----LREFASHLQQLDALNELT 345
QY 102 DEYN-----EKASAKQDILIRLDGKVKLNEAKSLTSSQSFNNASGKLLALDSQ 154
DB 346 EEHSKATQEWLEKQAKLEKLSAALQD---KKCLEEKNEIL-----QKLSQLEEH 393
QY 155 LT-----NDFSEKSSYFQS--QVDRIRKEAYAGAAGIVAGPFGI---IISYSIAAGVIEG 205
DB 394 LSQIQDNPPEKGEVLGDVLETLKQE-----AATLAANTQLOARVEMLETERGQGEA 448
QY 206 KLIFE-----LNRLKTVQNFSTSATVKQANKDIDAUKLATEIAA-IGEIK 254
DB 449 KLLAERGHFEKQKLSLITDQSSISNLS---QAKEELEQASQAHGARLTAQVASLT 504
QY 255 TETETTRFYDDYDMLSLKGAAGK---XMINTCNEYQO-----RH 292
DB 505 SELTILNATIOQQQBELAGLQKQAKEQAQLAQTLOQQEQASQGLRH 551

RESULT 35
US-08-467-781-4
Sequence 4, Application US/08467781
Patent No. 5780596
GENERAL INFORMATION:
APPLICANT: TOUKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HUWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,781
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-781-4

Query Match 6.4%; Score 96.5; DB 1; Length 2101;
Best Local Similarity 22.6%; Pred. No. 4.1;
Matches 65; Conservative 50; Mismatches 107; Indels 65; Gaps 14;
QY 42 ETIKELSRFKQYSEASVLDGDIKVLMDSDQKYFEATQTVYWGCVVTTLLSAYILLF 101
DB 294 ETLKQCDLTKESQ-----MDRKINQLSENGDLSFK-----LREFASHLQQLDALNELT 345
QY 102 DEYN-----EKASAKDILIRIDGDKVKKLNEAQSLLTSSQSFNNASGKLALDSQ 154
DB 346 EHSKATQEWLEKQAEKELSAALQD---KKCLEEKNEIL-----QKLSQLEEH 393
QY 155 LT-----NDFSEKSSYFQS--QVDRIRKEAYAGAAAGIVAGPGL---IISYSIAAGVIEG 205
DB 394 LSQLDNPPQEKGEVLDVQLQLETLKQ-----AATLAANTQLQARVEMLETERGGQQA 448
QY 206 KLIP-----LNRLKTVQNFFTSLSATVKQAKDIDAAKLKLATEIAA-IGEIK 254
DB 449 KLLAERGHFEKQQLSSLLTDLQSSISNLS-----QAKEELESQAQHGARLTQAVASLT 504
QY 255 TETETTRFYVDVDDMLSLKGAAK-----KMINTCNEYQQ-----RH 292
DB 505 SELTTLNATIQDQDELQAGLQKAQKQAEKQQLAQTLQOQEQASQGLRH 551

RESULT 36
US-08-195-487-4
Sequence 4, Application US/08/95487
Patent No. 5783403
GENERAL INFORMATION:
APPLICANT: TOUKATLY, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA HURWITZ & THIBEAULT
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,487
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US/07/901,701
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7100
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-195-487-4

Query Match 6.4%; Score 96.5; DB 1; Length 2101;
Best Local Similarity 22.6%; Pred. No. 4.1;
Matches 65; Conservative 50; Mismatches 107; Indels 65; Gaps 14;
QY 42 ETIKELSRFKQYSEASVLDGDIKVLMDSDQKYFEATQTVYWGCVVTTLLSAYILLF 101
DB 294 ETLKQCDLTKESQ-----MDRKINQLSENGDLSFK-----LREFASHLQQLDALNELT 345
QY 102 DEYN-----EKASAKDILIRIDGDKVKKLNEAQSLLTSSQSFNNASGKLALDSQ 154
DB 346 EHSKATQEWLEKQAEKELSAALQD---KKCLEEKNEIL-----QKLSQLEEH 393
QY 155 LT-----NDFSEKSSYFQS--QVDRIRKEAYAGAAAGIVAGPGL---IISYSIAAGVIEG 205
DB 394 LSQLDNPPQEKGEVLDVQLQLETLKQ-----AATLAANTQLQARVEMLETERGGQQA 448
QY 206 KLIP-----LNRLKTVQNFFTSLSATVKQAKDIDAAKLKLATEIAA-IGEIK 254
DB 449 KLLAERGHFEKQQLSSLLTDLQSSISNLS-----QAKEELESQAQHGARLTQAVASLT 504
QY 255 TETETTRFYVDVDDMLSLKGAAK-----KMINTCNEYQQ-----RH 292
DB 505 SELTTLNATIQDQDELQAGLQKAQKQAEKQQLAQTLQOQEQASQGLRH 551

RESULT 37
US-08-483-924-4
Sequence 4, Application US/08483924
Patent No. 5882876
GENERAL INFORMATION:
APPLICANT: TOUKATLY, GARY
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-924-4

Query Match 6.4%; Score 96.5; DB 2; Length 2101;
Best Local Similarity 22.6%; Pred. No. 4.1;

Matches		65; Conservative	50; Mismatches	107; Indels	65; Gaps	14;
QY	42	ETIKELSRFKQEYSQAEASVLVDGIDKIVLLMDSODKYFEATQTVEYGVVTVQLLSAYILLF	101			
Db	294	ETLKQCQDLKTEKSQ---- <td>345</td> <td></td> <td></td> <td></td>	345			
QY	102	DEYN-----EKKASAQKDILIRILDGVGYKLNEAKSILLTSQSFNNAASKILLALDSQ	154			
Db	346	EHSXATOWELEKAQLEKEKLSAALQ---KKCLEERNEIL-----OQKSLQLEE	393			
QY	155	LT-----NDFSPKSYFOS--QVDRIRKEAYAGAAIGIVAGPFGI-----IISYSIAAGVTEG	205			
Db	394	LSQLQDNPPQSEKGEVLGDVLQLETKE-----AATLAANNITQLOARVEMLETEREQQEA	448			
QY	206	KLIPE-----LNRLKTVCNFSTLSATVKQANKDIDAACKLKLATEIAA-IGRIK	254			
Db	449	KLLAERGHFESEKQCLASLTDLOSSTSNUS-----QAEELEQASQAHGARLTAQVASLT	504			
QY	255	TETETTRFYVDYDDLMLSLUNGAAK-----RVINTCNEYQQ-----RH	292			
Db	505	SLETTTNATIQQQOEELAGLKKQAKEQAOGLAQTLQOOEASOGLRH	551			

RESULT 38

US-09-452-294-1

; Sequence 1, Application US/09452294

Patent No. 6287790

; GENERAL INFORMATION:

APPLICANT: Lelievre, Sophie

APPLICANT: Bigsell, Mina

TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED

TITLE OF INVENTION: THERAPH AND DETECTION OF PROLIFERATIVE AND

TITLE OF INVENTION: DIFFERENTIATION DISORDERS

FILE REFERENCE: IB-1454- Sentence Submittal

Patent No. 6287790

: CURRENT APPLICATION NUMBER: IIS/09/452 294

: CURRENT FILING DATE: 1999-11-30

7. COMMENT FILING DATE: 1999-11-30
 : PRIOR APPLICATION NUMBER: 60/110,420

PRIOR FILING DATE: 1998-11-30

: NUMBER OF SEO ID NOS: 1

: SOFTWARE: Patent In Ver 3 1
: NUMBER OF SEQ IN NOS: 1

SEC ID NO 1

: LENGTH: 2101

REF ID: A66011

LIFE: FRI
ORGANISM: Homo sapiens

ORGANISM: HUMAN
TIS-09-452-294-1

Query Match 6.4%; Score 96.5; DB 3; Length 2101;
Best Local Similarity 22.6%; Pred. NO. 4.1;
Matches 65; Conservative 50; Mismatches 107; Indels 65; Gaps 14;

RESULT 39

; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-914-259-19

Query Match 6.3%; Score 96; DB 4; Length 975;
Best Local Similarity 21.1%; Pred. No. 1.4;
Matches 63; Conservative 56; Mismatches 118; Indels 62; Gaps 14;
QY 37 WKTFTDTIKELSRFKQYQSEASVLDGDIKVLMSQDKYFEATQTVVEMCGVVTQL--- 93
DB 376 WRA-GETVKAESQINMEDMEAS-----TPNLEVEAAQTAAERALAAQRTAL 422
QY 94 --LSAVILLFDEYNEK-KASAKDILIRLDGQVKLNEAKSLTSSQSFNNASGKLLA 150
DB 423 ANMSASVAV-----NEQARLATECERLYQQDDXDBEINO-----QSQYABQLKEQWNE 471
QY 151 LBSQLTNDSEKSSYFQSDVRIRKAYAG--AAAGIVAGPGLIISYSIAAGVIEGKL- 207
DB 472 QBELIANARREYET-LQSEMARITQENESAKBEVKEVLQALBELAVNYDQKSQEIDNKNK 530
QY 208 -IPELANRLKTVONFPTSLSATVQKANKDIDDAKLAKEIAA-----IGHI----- 253
DB 531 DIDALNEELQCKQSVENAASTELQQL-KWMSHQKRGITMTNLRDLGEVQQAIPGE 589
QY 254 -----KTETETTRFYVDYD--DLMSLLKGAAGKMINTCN--EYQQRHGKTLFE 299
DB 590 SSIDLKMSALAGTDASKVEEDFTMARLF-SMKTEAKNAQRCSNMETQADSNKKISS 648

RESULT 41

US-09-562-737-46
; Sequence 46, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
; APPLICANT: Herz, Joachim
; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 1024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Sequence
US-09-562-737-46

Query Match 6.3%; Score 95.5; DB 4; Length 1024;
Best Local Similarity 21.8%; Pred. No. 1.7;
Matches 74; Conservative 53; Mismatches 122; Indels 91; Gaps 16;
QY 1 MTSIFAQ-----TVEWKSATETADGALDLYNKYLDQVTPWKTDETIKELSRFKQYBS 55
DB 485 MPTLSAQALVGTINTSSQWQCAQATLDDYD-----TLPLQDDASKAWRKNEDES 539
QY 56 QEA-----SVLVGDIKVLMSQDKYFEATQTVVEMCGVVTQLLSAVILLFDEYNEKA 109
DB 540 KHEIHSFVDIATAGVNLVAGDPH-----ETDYAVGCIVTITSS-----NLTKM 587
QY 110 SAQKDILIRLD-----GVKKLNEAKSLTSSQ-----SFNNA 144
DB 588 SRGVKLLALLEGGSGMPLQAAKGLNGAVSELLRSQQPASBPORLLOAAGNVGSA 647

QY 145 SGKLLALDSQLTNDSEKSSYFQSDVRIRKAYAGAAAGIVAGPGLIISY-SIA----- 199
DB 648 SCELL---QQITE--SDTDPHQFVALMQLAKAVWSAAA-----LVLYKSVQORTE 694
QY 200 -AGVIEGKLIPELNRLKTVONFPTSLSATVQKANKDIDDAKLAKEIAA 250
DB 695 DAGLQTVIAADTQCALSTSQEVACTKVAPFISSPVCCQ--EQGVEAGRLVAKHVEGCVSA 753
QY 251 GEIKTETETTRFYVDYD--DLMSLLKGAAGKMINTCNEYQQ 290
DB 754 SQIATE-----DGQLLRKVGAATAVATLALNELLQ 783

RESULT 42

US-09-157-257-4
; Sequence 4, Application US/09157257
; Patent No. 6375954
; GENERAL INFORMATION:
; APPLICANT: DUTTA, Sukanta K.
; APPLICANT: BISWAS, Biswajit
; APPLICANT: VENUGOPALU, Ramesh
; TITLE OF INVENTION: A SIZE-VARIABLE STRAIN-SPECIFIC PROTECTIVE ANTIGEN FOR
; FILE REFERENCE: 8172-9016
; CURRENT APPLICATION NUMBER: US/09/157,257
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 60/059,252
; EARLIER FILING DATE: 1997-09-18
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 849
; TYPE: PRT
; ORGANISM: Ehrlichia risticii
US-09-157-257-4

Query Match 6.3%; Score 95; DB 4; Length 849;
Best Local Similarity 24.1%; Pred. No. 1.4;
Matches 76; Conservative 42; Mismatches 119; Indels 78; Gaps 16;
QY 41 DETIKELSRFKQYQSEASVLDGDIK--VLLMSQDKYFEATQTVVEMC--GVVTQLLSA 96
DB 300 DQTAKEVLK-----DSTAKDILKDTNAAAVLNKSTAKELTNTQTAKEVLTGDSKEVLKE 354
QY 97 YILLFDEYNEKASAKDILIRLDGQVKLNE-----AQKSLTSSQSFNNASGKL 148
DB 355 -ILTCDFKEAVTGDGDKILKGLTDTGKFKELIBSTGDKILKDLTDS-----TGKF 407
QY 149 -----LALDSQLTNDSEKSSYFQSDVRIRKAYAGAAAGIVAGP--GLIISYSTAAGV 202
DB 408 KELIEVLVRLNKIKELTDTNTGNFKGLVEGAGKDE---AKAVLTDERFKGLFDDKTIAGYV 454
QY 203 IE-----GK-----LIPELNNRL-----KTQVNFPTLSA-----T 228
DB 465 KEILTSEKPKLPESAGKTKVKKELLIDKFKQLFEDDTKASHVKEILTNDTAKEILTDQT 524
QY 229 VKANKDIDAAKLAKEIAAIGEIKTETETTRFYVDYDDMLMSLLKGAAGK-----KMINTC 285
DB 525 AKEVLKDSAKELIKDNTNAAALLKDSAKEVLK-----SDKFKDAITGAGKDALKEILTC 579
QY 286 NEYQQR---HGKKTLL 297
DB 580 DKFKEAVTGNKDIIL 594

RESULT 43

US-08-317-450B-15
; Sequence 15, Application US/08317450B
; Patent No. 5660982
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Kallunki, Pekka
; APPLICANT: Pyke, Charles

;; TITLE OF INVENTION: Laminin Chains: Diagnostic and
;; TITLE OF INVENTION: Therapeutic Use
;; NUMBER OF SEQUENCES: 19
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BANNER & ALLEGRETTI, LTD.
;; STREET: Ten South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/317,450B
;; FILING DATE: 04-OCT-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Chao, Mark
;; REGISTRATION NUMBER: 37,293
;; REFERENCE/DOCKET NUMBER: 94,778
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312-715-1000
;; TELEFAX: 312-715-1234
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1111 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-08-317-450B-15

Query Match 6.3%; Score 95; DB 1; Length 1111;

Best Local Similarity 17.9%; Pred. No. 2.1; Indels 72; Gaps 9;

Matches 48; Conservative 51; Mismatches 97; Indels 72; Gaps 9;

QY 10 VEVVKSATADGALDLYNKYLQVLPWKTFDETIKELSRFKQEQYSQ-----EAS 59
DB 868 VEEAKRIKQKADSLSSLVTRHMD-----FKTKQNLGNWKEAQQLONGKSGREKSD 921
QY 60 VLVGDIKVLMDSDQKYFEATQTVYEWGVVTTOLLSAYILLFD-EYNEKKASAK-----113
DB 922 QLLSRANLAKSRAQALSMGNATFYEVESILKNLRE-----FDLQVDRNRKAAEAEAMKRL 976
QY 114 DILIRIIDDGVKKLNEAKSL--LTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQV 170
DB 977 SYISQKVSASDKTQQAERALGSAADAQRAKNGAGEALEISSEI-----EQEI 1025
QY 171 DRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVK 230
DB 1026 GSLNLEAV-TADGALAMEKGL-----ASLKSEMR 1054
QY 231 QANKDIDAAKUKLATEIAAIGEIKTETE 258
DB 1055 EVEGELERKELEFDTNMDAVQMVITEAQ 1082

RESULT 44

US-08-800-593-15

Sequence 15, Application US/08800593

Patent No. 6143505

GENERAL INFORMATION:

APPLICANT: Tryggvason, Karl

APPLICANT: Kallunki, Pekka

APPLICANT: Pyke, Charles

TITLE OF INVENTION: Laminin Chains: Diagnostic and

TITLE OF INVENTION: Therapeutic Use

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

STREET: 300 South Wacker Drive

;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606
;;
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/800,593
;; FILING DATE: 18-FEB-1997
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/317,450
;; FILING DATE: 04-OCT-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Chao, Mark
;; REGISTRATION NUMBER: 37,293
;; REFERENCE/DOCKET NUMBER: 94,778-B
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312-913-0001
;; TELEFAX: 312-913-0002
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1111 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
US-08-800-593-15

Query Match 6.3%; Score 95; DB 3; Length 1111;

Best Local Similarity 17.9%; Pred. No. 2.1;

Matches 48; Conservative 51; Mismatches 97; Indels 72; Gaps 9;

QY 10 VEVVKSATADGALDLYNKYLQVLPWKTFDETIKELSRFKQEQYSQ-----EAS 59
DB 868 VEEAKRIKQKADSLSSLVTRHMD-----FKTKQNLGNWKEAQQLONGKSGREKSD 921
QY 60 VLVGDIKVLMDSDQKYFEATQTVYEWGVVTTOLLSAYILLFD-EYNEKKASAK-----113
DB 922 QLLSRANLAKSRAQALSMGNATFYEVESILKNLRE-----FDLQVDRNRKAAEAEAMKRL 976
QY 114 DILIRIIDDGVKKLNEAKSL--LTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSQV 170
DB 977 SYISQKVSASDKTQQAERALGSAADAQRAKNGAGEALEISSEI-----EQEI 1025
QY 171 DRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVK 230
DB 1026 GSLNLEAV-TADGALAMEKGL-----ASLKSEMR 1054
QY 231 QANKDIDAAKUKLATEIAAIGEIKTETE 258
DB 1055 EVEGELERKELEFDTNMDAVQMVITEAQ 1082

RESULT 45

US-08-317-450B-13

Sequence 13, Application US/08317450B

Patent No. 5660982

GENERAL INFORMATION:

APPLICANT: Tryggvason, Karl

APPLICANT: Kallunki, Pekka

APPLICANT: Pyke, Charles

TITLE OF INVENTION: Laminin Chains: Diagnostic and

TITLE OF INVENTION: Therapeutic Use

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: BANNER & ALLEGRETTI, LTD.

STREET: Ten South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,450B
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1193 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-450B-13

Query Match 6.3%; Score 95; DB 1; Length 1193;
Best Local Similarity 17.9%; Pred. No. 2.4;
Matches 48; Conservative 51; Mismatches 97; Indels 72; Gaps 9;

QY 10 VEVKSAIETADGALDLYNKYLQVVPKTFDTETIKELSRFKQEQYSQ-----EAS 59
DB 868 VEEAKRIKQADSLSSLVTRHDE-----FKRTQKLNKWKKEEAQQLQNGKSGREKSD 921
QY 60 VLVGDIKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFD-EYNEKKASAOK-----113
DB 922 QLLSRANLAKSAQAELSNGMNAIFYEVESILKNLRE-----FDLQVDNRKAEAEAMKRL 976
QY 114 DILIRILDGVKLNKAEQKSL---LTSSQSFNNSGKLLALDSQLTNDPSEKSSYFQSOV 170
DB 977 SVYSQKVSASDQKQAEALGSAADAQRAKNGAGEALEISSEI-----EQEI 1025
QY 171 DRIRKAYAGAAGVAGPFGGLIISYSIAAGVIEGKLIPELNNRLKTVQNFTSLSATVK 230
DB 1026 GSLNLEANY-TADGALAMEKGL-----ASLKSEMR 1054
QY 231 QANKDIDAAKKLATEIAAIGIKTETE 258
DB 1055 EVEGELERKELEFDTNMDAVQMVITEAQ 1082

RESULT 46

US-08-800-593-13
Sequence 13, Application US/08800593
Patent No. 6143505
GENERAL INFORMATION:
APPLICANT: Tryggvason, Karl
APPLICANT: Kallunki, Pekka
APPLICANT: Pyke, Charles
TITLE OF INVENTION: Laminin Chains: Diagnostic and
TITLE OF INVENTION: Therapeutic Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,593
FILING DATE: 18-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,450
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chao, Mark
REGISTRATION NUMBER: 37,293
REFERENCE/DOCKET NUMBER: 94,778-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1193 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-800-593-13

Query Match 6.3%; Score 95; DB 3; Length 1193;
Best Local Similarity 17.9%; Pred. No. 2.4;
Matches 48; Conservative 51; Mismatches 97; Indels 72; Gaps 9;

QY 10 VEVKSAIETADGALDLYNKYLQVVPKTFDTETIKELSRFKQEQYSQ-----EAS 59
DB 868 VEEAKRIKQADSLSSLVTRHDE-----FKRTQKLNKWKKEEAQQLQNGKSGREKSD 921
QY 60 VLVGDIKVLMSQDKYFEATQTVVWCGVVTQLLSAYILLFD-EYNEKKASAOK-----113
DB 922 QLLSRANLAKSAQAELSNGMNAIFYEVESILKNLRE-----FDLQVDNRKAEAEAMKRL 976
QY 114 DILIRILDGVKLNKAEQKSL---LTSSQSFNNSGKLLALDSQLTNDPSEKSSYFQSOV 170
DB 977 SVYSQKVSASDQKQAEALGSAADAQRAKNGAGEALEISSEI-----EQEI 1025
QY 171 DRIRKAYAGAAGVAGPFGGLIISYSIAAGVIEGKLIPELNNRLKTVQNFTSLSATVK 230
DB 1026 GSLNLEANY-TADGALAMEKGL-----ASLKSEMR 1054
QY 231 QANKDIDAAKKLATEIAAIGIKTETE 258
DB 1055 EVEGELERKELEFDTNMDAVQMVITEAQ 1082

RESULT 47

US-08-714-741-32
Sequence 32, Application US/08714741
Patent No. 6500613
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: McDaniel, Larry S.
APPLICANT: Swiatlo, Edwin
APPLICANT: Yother, Janet
APPLICANT: Crain, Marilyn J.
APPLICANT: Hollingshead, Susan
APPLICANT: Tart, Rebecca
APPLICANT: Brooks-Walter, Alexis
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
TITLE OF INVENTION: PORTIONS AND PRODUCTS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.
ZIP: 10036

```
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/714,741
;; FILING DATE: 16-SEP-1996
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Frommer Esq., William S.
;; REGISTRATION NUMBER: 25,506
;; REFERENCE/DOCKET NUMBER: 454312-2460
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 840-3333
;; TELEFAX: (212) 840-0712
;; INFORMATION FOR SEQ ID NO: 32:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 8991 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: amino acid
;; US-08-714-741-32

Query Match      6.3%; Score 95; DB 4; Length 8991;
Best Local Similarity 19.7%; Pred. No. 54;
Matches 63; Conservative 50; Mismatches 116; Indels 90; Gaps 11;

QY 6 AEQTVVEVKSATADGALDLYNKYLDQVTPKTFDTIKELSRFKQYSEASVVGDI 65
Db 2700 AKDAKAKKAVEDAQKALD-----DAKAAQKKYDE-----DQKTEEKAAL----- 2741

QY 66 KVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLDFEYNEKKAQKDIILRLDDGVK 125
Db 2742 EKAASEMDKAAVAQVAY-----KSLTSSQSFNNASGKLLALDSQ-----TNDFSKSY 165

QY 126 KLINEAQ-----KSLTSSQSFNNASGKLLALDSQ-----TNDFSKSY 165
Db 2785 REBEAKTKFNTVRAMVPEPEQAEATKKSEBAKQKAPELTKLEBAKALEAEKXATE 2844

QY 166 FQSOVDRIKAEYAGAAAGVAGPFGLLIISYSIAAGVIEGKLIPELNNRLKTVQNF 225
Db 2845 AKQKVDAMKKKMLTSLASVA-----ILGAGIVASQ-----PTLVRAEESP 2886

QY 226 SATVQKANKIDAA-----KLKATETIAAIGELTKTETTRFYVDYDMLSL 273
Db 2887 VASQSKAEKDYDAVKKSEAAYEBAKALEAEKVAQKKYEDDQKK-----TESKAE 2941

QY 274 LKGAACKKWINTCNEYQQRH 292
Db 2942 EKEASEAIAKATEEVQAY 2960

RESULT 48
US-09-934-899-8
; Sequence 8, Application US/09934899
; Patent No. 6537786
; GENERAL INFORMATION:
; APPLICANT: Wang, Siqun
; APPLICANT: Wang, Tao
; APPLICANT: Kofias, Mattheos
; APPLICANT: Odom, J. Martin
; APPLICANT: Ye, Rick
; TITLE OF INVENTION: Genes encoding exopolysaccharide production
; FILE REFERENCE: C11633 US NA
; CURRENT APPLICATION NUMBER: US/09/934,899
; CURRENT FILING DATE: 2001-08-22
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 779
; TYPE: PRT

; ORGANISM: Methylomonas 16a
US-09-934-899-8

Query Match      6.2%; Score 94.5; DB 4; Length 779;
Best Local Similarity 18.2%; Pred. No. 1.4;
Matches 65; Conservative 66; Mismatches 133; Indels 93; Gaps 12;

QY 2 TSIFAEQTVVEVKSATET-----ADGALDLYNKYLDQVTPKTFDE 42
Db 228 TSLAAIETLQKAFSVKSVKSDNTLSVELKGRDEQLAKSVNDIASIYVATVNNES--- 284

QY 43 TIKELSRFKQYSEASVVGDIKVLMDSDQKYFEATQTVYEWCGVVTQLLSAYILLDF 102
Db 285 -----AEASQKLNFLSQLP-LYKENLEKAEQALSAYRQHGAVDLSAEAEILL--- 332

QY 103 EYNEKKAQAQKDIILRI-----LDDGVKLINEAQKSLTSSQSFNNASGKLLALDSQ----- 154
Db 333 ----KQASEMETLSIQLKQKYDEQSRLESEHPDMIATNAQIRRVSNKLALEKRIKDL 388

QY 155 -----LTNDFSEKSYFQSOVDRIKAEYAGAA-----AGIVAGPF-----GL 192
Db 359 KTOQNVYSLSRDVQVNTELYTSLNSAQEQRIAAAGSLGNSRIVDFAVVPEKYPKPG 448

QY 193 IISYSIAAGVIEGKLIPELNNRLKTVQNF-----FTLSATVQK-----A 232
Db 449 LLATAGLLIGISGLSALIFLRHSLQRHNDYFALLEYQVGLPFAAIPHSKKQRLARLLDQ 508

QY 233 NKDIDAAKL-----KLATETIAAIGELTKTETTRFYVDYDMLSL-LKGAACKKWINT 284
Db 509 GKERTAILVSHDPLDLSVESLGRITLNTLASDESKVIMVSSPAPGKSGKSFIST 565

RESULT 49
US-09-081-689-2
; Sequence 2, Application US/09081689
; Patent No. 6165992
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola G.
; APPLICANT: Zallacain, Madgalenaa
; APPLICANT: Throup, John
; APPLICANT: Biswas, Sanjoy
; TITLE OF INVENTION: Histidine Kinase
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,689
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,347
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GW10009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 442 amino acids
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2003, 13:50:27 ; Search time 31 Seconds
(without alignments)
1814.684 Million cell updates/sec

Title: US-09-993-292a-2

Perfect score: 1515

Sequence: 1 MTSIFABQTVVVKGAIFTA.....NEYQORHGKTLFEVFDVAS 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 573684 seqs, 184443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1515	100.0	305	10	US-09-993-292a-2
2	116	7.7	1023	11	US-09-884-696-5
3	112.5	7.4	1948	12	US-10-032-585-7611
4	106.5	7.0	1312	12	US-10-393-602-148
5	106	7.0	2835	10	US-09-885-535-4
6	104.5	6.9	3158	9	US-09-815-242-12611
7	102	6.7	1038	12	US-10-032-585-7776
8	101	6.7	435	11	US-09-978-309A-80
9	101	6.7	476	11	US-09-978-309A-77
10	100.5	6.6	946	15	US-10-156-761-14486
11	100	6.6	722	9	US-09-815-242-10796
12	100	6.6	2871	15	US-10-146-473-41
13	99.5	6.6	476	11	US-09-978-309A-79
14	99.5	6.6	631	11	US-09-978-309A-48
15	99.5	6.6	660	12	US-09-841-260-139

16	99.5	6.6	660	14	US-10-007-693-139	Sequence 139, App
17	99.5	6.6	1259	12	US-10-032-585-7150	Sequence 7120, App
18	98	6.5	1095	15	US-10-128-714-3039	Sequence 3039, App
19	98	6.5	1223	11	US-09-964-256A-2	Sequence 2, Appli
20	98	6.5	1223	12	US-10-146-733-44	Sequence 44, Appli
21	98	6.5	1277	15	US-10-128-714-8039	Sequence 8039, App
22	98	6.5	1648	15	US-10-157-031-38	Sequence 38, Appli
23	97.5	6.4	759	9	US-09-815-242-5060	Sequence 5060, App
24	97.5	6.4	759	11	US-09-975-713-170	Sequence 170, App
25	97.5	6.4	1938	11	US-10-171-311-164	Sequence 164, App
26	97.5	6.4	1945	11	US-09-927-597-2	Sequence 2, Appli
27	97.5	6.4	1972	12	US-10-341-434-103	Sequence 103, App
28	97.5	6.4	1972	15	US-10-171-311-162	Sequence 162, App
29	97.5	6.4	1979	11	US-09-927-597-4	Sequence 4, Appli
30	97.5	6.4	2189	12	US-10-172-503-2	Sequence 2, Appli
31	97.5	6.4	2285	10	US-09-932-183A-2	Sequence 2, Appli
32	97	6.4	961	12	US-10-080-608A-66	Sequence 66, Appli
33	97	6.4	961	12	US-10-080-608A-155	Sequence 155, App
34	97	6.4	6281	9	US-09-815-242-12996	Sequence 12996, A
35	96.5	6.4	265	10	US-09-987-107-21	Sequence 21, Appli
36	96.5	6.4	457	11	US-09-883-343A-6	Sequence 6, Appli
37	96.5	6.4	3562	12	US-10-341-434-109	Sequence 109, App
38	96	6.3	523	12	US-09-976-782-6	Sequence 6, Appli
39	96	6.3	975	12	US-10-080-608A-19	Sequence 19, Appli
40	96	6.3	975	12	US-10-370-685-108	Sequence 108, App
41	96	6.3	1710	12	US-10-032-585-7238	Sequence 7238, App
42	95.5	6.3	1024	15	US-10-211-962-46	Sequence 46, Appli
43	95	6.3	613	15	US-10-013-477-18	Sequence 18, Appli
44	95	6.3	613	15	US-10-013-477-26	Sequence 26, Appli
45	95	6.3	613	15	US-10-207-655-204	Sequence 204, Appli
46	95	6.3	1111	9	US-09-756-071B-15	Sequence 15, Appli
47	95	6.3	1111	15	US-10-227-738-15	Sequence 15, Appli
48	95	6.3	1193	9	US-09-756-071B-13	Sequence 13, Appli
49	95	6.3	1193	15	US-10-227-738-13	Sequence 13, Appli
50	94.5	6.2	329	9	US-09-925-302-490	Sequence 490, App
51	94.5	6.2	779	10	US-09-934-899-8	Sequence 8, Appli
52	94.5	6.2	779	10	US-09-934-868-28	Sequence 28, Appli
53	94.5	6.2	779	12	US-10-353-457-8	Sequence 8, Appli
54	94.5	6.2	779	12	US-10-353-456-8	Sequence 8, Appli
55	94.5	6.2	2434	9	US-09-815-242-5835	Sequence 5835, App
56	94.5	6.2	5795	9	US-09-815-242-12610	Sequence 12610, A
57	94	6.2	1193	12	US-10-053-662A-31	Sequence 31, Appli
58	94	6.2	1193	15	US-10-171-311-115	Sequence 115, App
59	94	6.2	1881	12	US-10-032-585-7646	Sequence 7646, App
60	94	6.2	2186	9	US-09-815-242-12913	Sequence 12913, A
61	93.5	6.2	442	9	US-09-737-068-2	Sequence 2, Appli
62	93.5	6.2	647	9	US-09-841-132-436	Sequence 436, App
63	93.5	6.2	928	12	US-10-080-608A-23	Sequence 23, Appli
64	93.5	6.2	928	12	US-10-370-685-112	Sequence 112, App
65	93.5	6.2	1111	12	US-10-080-608A-28	Sequence 28, Appli
66	93.5	6.2	1111	12	US-10-370-685-117	Sequence 117, App
67	93	6.1	996	9	US-09-815-242-5231	Sequence 5231, App
68	93	6.1	1009	9	US-09-815-242-12141	Sequence 12141, A
69	92.5	6.1	281	12	US-10-080-608A-63	Sequence 63, Appli
70	92.5	6.1	281	12	US-10-370-685-152	Sequence 152, App
71	92.5	6.1	284	12	US-10-080-608A-48	Sequence 48, Appli
72	92.5	6.1	284	12	US-10-080-608A-50	Sequence 50, Appli
73	92.5	6.1	284	12	US-10-370-685-137	Sequence 137, App
74	92.5	6.1	284	12	US-10-370-685-139	Sequence 139, App
75	92.5	6.1	321	10	US-09-967-347-12	Sequence 12, Appli
76	92.5	6.1	1610	15	US-10-155-533-9	Sequence 9, Appli
77	92	6.1	722	9	US-09-815-242-12888	Sequence 12888, A
78	92	6.1	935	12	US-10-080-608A-25	Sequence 25, Appli
79	92	6.1	935	12	US-10-370-685-114	Sequence 114, App
80	92	6.1	991	9	US-09-815-242-5803	Sequence 5803, App
81	92	6.1	1184	9	US-09-815-242-5329	Sequence 5229, App
82	92	6.1	1188	9	US-09-815-242-12125	Sequence 12125, A
83	92	6.1	2086	9	US-09-815-242-5639	Sequence 5639, App
84	91.5	6.0	243	12	US-09-969-748C-108	Sequence 108, App
85	91.5	6.0	284	12	US-10-080-608A-55	Sequence 55, Appli
86	91.5	6.0	284	12	US-10-370-685-144	Sequence 144, App
87	91.5	6.0	516	12	US-10-032-595-7147	Sequence 7147, App
88	91.5	6.0	746	10	US-09-982-107-4	Sequence 4, Appli

Sequence 1, Appli
Sequence 59, Appl
Sequence 124, App
Sequence 65, Appl
Sequence 9, Appl
Sequence 10, Appl
Sequence 967-347-10
Sequence 7537, Ap
Sequence 7431, Ap
Sequence 11, Appl
Sequence 4, Appl
Sequence 2, Appl
Sequence 100, App

ALIGNMENTS

RESULT 1
US-09-993-292a-2
; Sequence 2, Application US/09993292A
; Patent No. US20020146430A1
; GENERAL INFORMATION:
; APPLICANT: James E. Galen
; TITLE OF INVENTION: USE OF CLY A HEMOLYSIN FOR EXCRETION OF
; FILE OF INVENTION: PROTEINS
; FILE REFERENCE: UOFMD.007A
; CURRENT APPLICATION NUMBER: US/09/993,292A
; PRIOR FILING DATE: 2001-11-23
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Salmonella Typhi
US-09-993-292a-2

Query Match 100.0%; Score 1515; DB 10; Length 305;
Best Local Similarity 100.0%; Pred. No. 4.9e-129;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLDQVVPKTFDETIKLSRFKQYSEASV 60
DB 1 MTSIFAEQTVVVKSAIETADGALDLYNKYLDQVVPKTFDETIKLSRFKQYSEASV 60
QY 61 LVGDIKVLNDSQDKYFEATQTVYEWGVTQLLSAYILLFDEYNEKKSAAQKDIILIRIL 120
DB 61 LVGDIKVLNDSQDKYFEATQTVYEWGVTQLLSAYILLFDEYNEKKSAAQKDIILIRIL 120
QY 121 DDGVKCLNEAKSLTSSQFNNSAGKLLDLSQTLNDFSEKSYFQSQVDRIKREAYAG 180
DB 121 DDGVKCLNEAKSLTSSQFNNSAGKLLDLSQTLNDFSEKSYFQSQVDRIKREAYAG 180
QY 181 AAGIVAGPGLIISYIAAGVIEGKLIPELNNRLKTQVNFSTLSATVKQANKDIDAAK 240
DB 181 AAGIVAGPGLIISYIAAGVIEGKLIPELNNRLKTQVNFSTLSATVKQANKDIDAAK 240
QY 241 LKLAATAAIGEIKTETTRFYVDYDMLSLKGAKKMINTCNEYQORHGKKTLPFV 300
DB 241 LKLAATAAIGEIKTETTRFYVDYDMLSLKGAKKMINTCNEYQORHGKKTLPFV 300
QY 301 PDVAS 305
DB 301 PDVAS 305

RESULT 2
US-09-884-696-5
; Sequence 5, Application US/09884696
; Publication No. US20030035809A1
; GENERAL INFORMATION:

; APPLICANT: GEORGE, LISLE W
; APPLICANT: ANGELOS, JOHN A
; APPLICANT: HESS, JOHN F
; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES
; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA
; FILE OF INVENTION: BOVIS INFECTIONS
; FILE REFERENCE: 481.06
; CURRENT APPLICATION NUMBER: US/09/884,696
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-884-696-5

Query Match 7.7%; Score 116; DB 11; Length 1023;
Best Local Similarity 20.5%; Pred. No. 0.14;
Matches 69; Conservative 59; Mismatches 111; Indels 98; Gaps 11;

QY 2 TSIFAEQTVVVK-----SAIETADGALDLYNKYLDQVVPKTFDETIKLSRFKQE 53
DB 155 TALSMKIDELIKKQSGNVSSSELAKASIELINQLVDTA-----ASLNNVNSFQQ 207
QY 54 YSQEASVL-----VGDIKVLNDSQDKYFEATQTVYEWGVTQLLSAYILL-FDEY 104
DB 208 LNKLGSLVSLNTKHLNGVN-KLQNLPLNDNIGAGLDTV---SGILSAISAFILSNADAD 263
QY 105 NEKKSAAQKDIILIRLDD---GVKKLNEAKSLTSSQFNNSAGKLLDLSQTLNDFSE 161
DB 264 TGTAAAGVELTTKVLGNVGKISQYIIAQAAAGLST----- 302
QY 162 KSSYFQSQVDRIKREAYAGVAGPGLIIS-----YSIAAGVIEGKLIPELNNRLKT 217
DB 303 -----AAAAGLIASVVTILATISPLSFLSTADKFRANKIEEYSQRFKK 344
QY 218 VQNFSTLSATVKQANKDIDAAKLAETAAIGEIKTETTRFYVDYDMLSLKGA 277
DB 345 LGYDGLSLLAAFHKETGAIDASTRISTVLASVSSGISAATSLV---GAPVSALVGA 400
QY 278 -----AKQMINTCNEYQORHGK 294
DB 401 VTGIISGILEASKQAMPEHVASQADVAEWEKKGK 437

RESULT 3
US-10-032-585-7611
; Sequence 7611, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussy
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7611
; LENGTH: 1948
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7611

Query Match 7.4%; Score 112.5; DB 12; Length 1948;
Best Local Similarity 22.2%; Pred. No. 0.7;
Matches 58; Conservative 40; Mismatches 90; Indels 73; Gaps 10;
QY 38 KTFDETIKLSRFKQYSEASVVLGDIKVLNDSQDKYFEATQ-----TVYEWGVTQ 92

Db 1444 KTLKSTREELNGSKTEILR-----LKALLRESEDELYQVQKQNYKTSVHDYEDQLAQ 1495
Qy 93 L-----LSAVILLDFEYNEKASAKDILI--RIILDDGVKKLNEAQS 133
Db 1496 LKVKHETLLSRKNDINSLIYKRSDEYYKLELAESATAISXHEQATKEMKESRSQ 1555
Qy 134 LITSQSNNASGKLLALDSOLTNDPSEKSYFOSQVDRIRKEAYAGAAAGIVAGPPGLI 193
Db 1556 LLLVREE-----LRTQILIKDFRIKVENLEATIEE-----1586
Qy 194 ISYSTAAGVIRKLIPE--LNNRLKTQV--FFTSLSATVQKANKDIDAALKLATETAAL 250
Db 1587 KKHLDANKKELKIQDKLNLVHLNFKFNKELNEKLEIKLNRDLB-----FKIDI---1638
Qy 251 GEIKTETETTFRYDDYDML 271
Db 1639 -ETKLIKENKKLQDYEDVLL 1658

RESULT 4

US-10-393-602-148
; Sequence 148, Application US/10393602
; Publication No. US20030170714A1
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESS: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/393,602
; FILING DATE: 19-Mar-2003
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/POCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
; SEQUENCE DESCRIPTION: SEQ ID NO: 148:

US-10-393-602-148
Query Match 7.0%; Score 106.5; DB 12; Length 1312;
Best Local Similarity 18.9%; Pred. No. 1.4;
Matches 75; Conservative 54; Mismatches 98; Indels 169; Gaps 16;
Qy 10 VEVVKSATETAGALDLYNKLVDVLPWKTFDETIK-----ELSRFKQSY 54
Db 502 MEVISLQNEKAD--LDRTLKLDQEMEQLNHHTTTTQMEMLTKDKADKDEQIRKISRH 559

Qy 55 SOEASVVGDIKVLMLDSQDKYFEATQTYEWCVVTTOLLSAYILLDFEYNEKASAKD 114
Db 560 SDELTSLLG-----YFPNKKQLEDWLHKSXK-----EINQTR-----591
Qy 115 ILIRILDDGVKKLNEAQSLLTSQSNNASGKLLALDSOLTNDPSEK-----SSYFQS 168
Db 592 -----DRLAKLN--KELASSEQNKHNNELKRREQLSS-YEDKLFVCGSQDFES 640
Qy 169 QVDRIRKE-----AVAGAAAGIVAGPF-----190
Db 641 LIDLRLKEEIEKSKQKRAMLAGATA--VYSQFTITQLTNDNQSCCPVCQRFQTEAELOQV 698
Qy 191 -----GLI-ISYSTAAGVIRKLIPELNNRLKT 217
Db 699 SLOSKLRLAPDKLSTESSELKKEKRRDEMLGLVPMRQSIID--LKEKEIPELENKLQ 756
Qy 218 VQNFFTLSATVQKANKDIDAALKLATETAIAAIGIKTETETT-----RIFYDY 266
Db 757 V-----NRDIQRLKNDIEFQETILGTIMPEESAKVCLTDVIMERFQWEL 802
Qy 267 DDLMLSLKGAAK-----KMINTCNEYQQRHGKK 295
Db 803 KQVERKIAQAAKLOGIDLDLDRTVQVQVQKQKQKH 838

RESULT 5

US-09-885-535-4
; Sequence 4, Application US/09885535
; Patent No. US20020104105A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Heichman, Karen
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-266-II
; CURRENT APPLICATION NUMBER: US/09/885,535
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/213,245
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 2835
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-885-535-4

Query Match 7.0%; Score 106; DB 10; Length 2835;
Best Local Similarity 19.4%; Pred. No. 4.7;
Matches 74; Conservative 76; Mismatches 122; Indels 110; Gaps 17;
Qy 7 EOTVEVVK-----SAIETAD-----GALDLYNKLVDV-IPWKT 39
Db 807 KSTIVEVIRKEGKIATPAEDKVKILKQLSLDSRWEALLNKAEITNEQLSGISVVAQ 866
Qy 40 FDETIELSRFK-----QEYSQEASVL---VGDIKVLMLDSQDKYFEATQTYEWM 86
Db 867 FHETLEPLNEWLTTEIKRLVNCPEIGTQASKLEEQIAQHKVLQEDILLRKQNVDAQALL 926
Qy 87 CGVVTTOLLSAYILLDFEYNEKASAKDII-----LIRILDDGV--KKLNEAQSLLTS 137
Db 927 LELLKQTTGDEVLITQDKLEAIKARYKIDITKSTVAKTLEQALQALRLHSTHELCT- 985
Qy 138 SOSFNNASGKLLALDSOLTNDPSEKSSYFOSQVDRIRKEAYAGAAAGIVAGPPGLI 197
Db 986 --WLDKVEVELLSYETQVLK--GEASQAQMRPKELKEAKNKA-----LLDSL 1032
Qy 198 -IAAGVIE-----CKLIPENLNNRLKTQNFSTLSATVQKANKDIDA-----238
Db 1033 EVSSALLELVPRWAREGLEKVAEDNERVRLV-----SDTITQKVEIDAILRSQOF 1085
Qy 239 -----AKLKLATE-----IAAIGEIKTETETTFRYVD-----YDDLMLSLKGA 278

Db 1086 DQADAELEWITEKLSGLRLEQDQTSALQVQKFTWELRHKDIIDDLVK-SG 1144
QY 279 KKMINTCNEVQQRHGKKTTFV 300
Db 1145 HKIMTACSEBEKQSMKKKLDKV 1166

RESULT 6

US-09-815-242-12611
; Sequence 12611, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITEA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12611
; LENGTH: 3158
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12611

Query Match 6.9%; Score 104.5; DB 9; Length 3158;
Best Local Similarity 24.1%; Pred. No. 7.4;
Matches 67; Conservative 26; Mismatches 114; Indels 71; Gaps 10;

QY 43 TIKELSRFKQYSQSEASVGLVDIKVLLMDSQDKYFEA-----TQTVYEWCGVWQ 92
Db 507 TLKQIQANSVQSVDFQMD-----QDKQAYNNAQAQIANGTPTPVLTPDTVTQ 561
QY 93 LLSAYILLFDEYN--EKKASQKDLIRILDGKVKLNEAQSLL-----135
Db 562 AVTTMNAQKDALNGDKLAQAKQDALNL--DTRLNQPQDALRNQINCAQALATVEQ 619
QY 136 --TSQSNNASGKLLALDSQLTDFSEKSYFQSVQVDRIRKEAYAGAAAGIVAGPGLI 193
Db 620 TKQNSQNVNTAWN--LKQGIANKQTVKASENTHDADQKQATYTNV-----665
QY 194 ISYIAAGVIRGKLIPELN-----NRLKTVQNFFTSL-----ATVKQANKDI-----236
Db 666 ---SQAGEIINQTNPTLPDEIRALTQVTDKXNGLNGEAKLATEKQNAKDAVNAMTHL 722
QY 237 -DAKKLKATIAAIGETKTETETTRFVYVDDMLSL 273
Db 723 NDAQKALQGQIDQSPETATVQVKQTATSLDQAMNQL 760

RESULT 7

US-10-032-585-7776
; Sequence 7776, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7776
; LENGTH: 1038
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7776

Query Match 6.7%; Score 102; DB 12; Length 1038;
Best Local Similarity 20.8%; Pred. No. 2.6;
Matches 64; Conservative 59; Mismatches 118; Indels 66; Gaps 13;

QY 7 EQTVVVKV---ALETADGALDLINKYLDQVIPKTFDETIKLSRFKQF-YSQEASVLV 62
Db 237 EETIQAALKSERNELTAKVSELEDYMKHSE-----VEEDVVMKQNDQFEQRIHELEAAIDT 291
QY 63 GDIKVLLMDSQDKYFEATQTVYEWCGVVTQLLSAYILLFDEYNKKASQKDLIRILD- 121
Db 292 LHQTEATIQQSQSFENTE-----LQKQLSTELDKQEMNELLASKNENLEMDLSEK 344
QY 122 -DGVKXLINEAOKSLTSSQSFNNASGKLLALDSQLTN--DFSEK-SSYFQSVQVDRIR-KE 176
Db 345 TDNLKELN---NKVLSQAQEIINLETKLDTLNSQFENNTDGNKLMKNLESQHKVQTQE 401
QY 177 AVAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNRLKTVQNFFTSLSATVKQANKDI 236
Db 402 AF-----IDELHHEQKTIIDNEY-----KAKIKDL 425
QY 237 DAAKKLKATIAAIGETKTETETTRFVYVDDML--SLKGAAGKQKMTNCFVQQR--- 291
Db 426 EVENALQSEETISRIRAKNSQVDPEAQHVEIDQLKQENALQKDNVKNLYNNFKELDKREVE 485
QY 292 HGKKTFL 298
Db 486 HAHQIAP 492

RESULT 8

US-09-978-309A-80
; Sequence 80, Application US/09978309A
; Publication No. US20030100490A1
; GENERAL INFORMATION:
; APPLICANT: Cruz, Tony
; APPLICANT: Pastrak, Aleksandra
; APPLICANT: Turley, Eva A.
; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to
; TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by
; TITLE OF INVENTION: Hyaladherin and Hyaluronans
; FILE REFERENCE: 033352-010
; CURRENT APPLICATION NUMBER: US/09/978,309A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 09/685,010
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: US 09/541,522
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/127,457
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80

RESULT 11

US-09-815-242-10796
; Sequence 10796, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10796

TYPE: PRT
ORGANISM: Enterococcus faecalis

US-09-815-242-10796

Query Match 6.6%; Score 100; DB 9; Length 722;

Best Local Similarity 19.5%; Pred. No. 2.3;

Matches 65; Conservative 52; Mismatches 109; Indels 108; Gaps 12;

QY 16 AIEATGAGLDLYNKYLDVIFPKTFDETIKELSPKQY-----SQE 57

DB 122 AVKGNDKIKSYDDLKGKTVAAKVGTSANFLEKNEKDYTIKNFDDATGLYKALENGE 181

QY 58 ASVLVDI:KVL-----LMDSDQ-----KYFEATQTVVEWCG 88

DB 182 ADAIVDDYPVLGYAVKNGKQLQVGDKETGSSYGFVAKGQNPelikKFNAGLKNLKDNG 241

QY 89 VTQLLSAYILLFPEYNEKASAKDILIRLDGKVLKNEAKSLTSSQS-----FN 142

DB 242 TYDKILNNYLGATGDTNTQDAGEQ-----MKKITPKKEKYVIASDSTFAPPEFQ 290

QY 143 NASGKLLALDSQLTNDFEKSYSY-----FQSQVDRIKKEAVAGAAAGIVAGPGLII 194

DB 291 NAQGDYGVGDVLDVKRAELQGFVEFKFICFSSAVQAVE-----SGQADGNVAG----- 340

QY 195 SYSIAAGVIEKGLPELNNRK-----TVQNFTSLSATVQKQAN---KDIDAAK-----LK 242

DB 341 -----NTITDRKAFDFSPYFSDSGIQIAVKGNDKIKSYDDLKGYKGVK 387

QY 243 LATEIA-AIGIKTETETTRFYVDYDDLMLSLK 275

DB 388 IGTSADFEKNNKKYDYSIKYLTDTDALYSALE 421

RESULT 12

US-10-146-473-41

; Sequence 41, Application US/10146473

; Publication No. US20030108888A1

; GENERAL INFORMATION:

APPLICANT: Scanlan, Matthew
APPLICANT: Gout, Ivan
APPLICANT: Stockert, Elisabeth
APPLICANT: Gure, Ali
APPLICANT: Chen, Yao-Tseng
APPLICANT: Old, Lloyd
TITLE OF INVENTION: Breast Cancer Antigens
FILE REFERENCE: L00461/70130(JRV)
CURRENT APPLICATION NUMBER: US/10/146,473
CURRENT FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: US 60/291,150
PRIOR FILING DATE: 2001-05-15
NUMBER OF SEQ ID NOS: 82
SOFTWARE: PatentIn version 3.0
SEQ ID NO 41
LENGTH: 2871
TYPE: PRT
ORGANISM: Homo sapiens
US-10-146-473-41

Query Match 6.6%; Score 100; DB 15; Length 2871;

Best Local Similarity 17.8%; Pred. No. 17;

Matches 60; Conservative 57; Mismatches 93; Indels 128; Gaps 12;

QY 25 DLYNKYLDVIFPKTFDETIKELSRFKQYEQEASVLVG-----DIKVLMLDSQDKYF 77

DB 1205 NLRNKYTEINITKT---TIKEISQKEDDSKLNRLNQLDRLSRENRLKDEIVRLNDSIL 1261

QY 78 EATQ-----TVYEMCGVVTQLLSAYILLFPEYNEKASAO---KDILIRILDGKVK 125

DB 1262 QATEQRRRAENALQKACG-----SEIMQKQHLIELQVMQQRSEDNAR 1308

QY 126 ---KLENAQKSLTSSQSF-----NNASGKLLALDS----- 153

DB 1309 HQQSLEAAKTQDKNKEIERLKAEFQBEAKRWEYENELSKVRNNYDEEIIISLKNQPET 1368

QY 154 -----OLTNDFFSEKSVFQSOVDRIKKEAVAGAAAGIVAGPGLIIISYSIAAGVI 203

DB 1369 ENITKTHIHLTMQKEEDTGYRAQIDNLTRENS----- 1404

QY 204 EGKLIPELNNRLKTVQNFTSLSATVQKQANXDIDAAKL-----KLATEIAAIGEI 253

DB 1405 -----LSEIEIKRLKNTLTOTTENLRREVEDIQQKATGSEVSRKQQLLEVELRQVTQM 1457

QY 254 KTETTTTFYVDYDDLMLSLKGAQKQKQINTCNEYQOR 291

DB 1458 RTE-ESVRYKQSLDD-----AAKTQDKNKEIER 1485

RESULT 13

US-09-978-309A-79

; Sequence 79, Application US/09978309A

; Publication No. US20030100490A1

; GENERAL INFORMATION:

APPLICANT: Cruz, Tony

APPLICANT: Pastrak, Aleksandra

APPLICANT: Turley, Eva A.

TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to

TITLE OF INVENTION: Injury and Other Proliferating Cell Disorders Regulated by

TITLE OF INVENTION: Hyaladherin and Hyaluronans

FILE REFERENCE: 033352-010

CURRENT APPLICATION NUMBER: US/09/978,309A

CURRENT FILING DATE: 2002-04-04

PRIOR APPLICATION NUMBER: US 09/685,010

PRIOR FILING DATE: 2000-10-05

PRIOR APPLICATION NUMBER: US 09/541,522

PRIOR FILING DATE: 2000-04-03

PRIOR APPLICATION NUMBER: US 60/127,457

PRIOR FILING DATE: 1999-04-01

NUMBER OF SEQ ID NOS: 84

SOFTWARE: FastSeq for Windows Version 4.0

APPLICANT: Bhatia, Ajay
APPLICANT: Probst, Peter
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
FILE REFERENCE: 210121.515C2
CURRENT APPLICATION NUMBER: US/10/007,693
CURRENT FILING DATE: 2001-12-05
NUMBER OF SEQ ID NOS: 157
SEQ ID NO 139
LENGTH: 660
TYPE: PRT
ORGANISM: Chlamydia trachomatis
US-10-007-693-139

Query Match 6.6%; Score 99.5; DB 14; Length 660;
Best Local Similarity 22.0%; Pred. No. 2.3;
Matches 70; Conservative 47; Mismatches 114; Indels 87; Gaps 16;
QY 8 QTVFVKSATETADGALDLYNKYL-----DQV-----IPWKT 40
DB 155 EVVNIKALEAQKDTDKLKLVTLQNKSLTEVLKTTDSADQIPAINSLKNSA 214
QY 41 DETIKELSRKQBSQASVILVDIKVLLMDSQD--KYFATQTVVWCGVVTOLLSAYI 98
DB 215 DQIKDLER--QNTSYEA-VLTNAGEVIKASSEAGIKLQALQSIYD-AGDQSQ---AAV 267
QY 99 LFFDEYNEKKAQAKDILIRILDGVKKLNEAQKSLTSSQSFNNASGKLLALDLSQTLND 158
DB 268 LQAQNNSPDNIAATKELIDAAETKVNELKQHTGL-----TDSPLVK 311
QY 159 FSEKSSVFSQSDVRKEAVAGAAAGIYVAGPGLIISYTAAGVIEGK-----LIPE 210
DB 312 ABEQISQAQKDIQIKP---SGSDPIV-GPSSAASAGSAGALKSSNNSGRISLLDD 367
QY 211 LNNELKTV--QNFETLSATVKQAN-----KQIDAQKLKATE-----IAAIGEIKTET 257
DB 368 VDNEWALALQGF---RSMIEQFVNVPATKELQAEALQAMSDQLVAGDELPAEI 423
QY 258 ETRFVYVDYDMLSLK 275
DB 424 QAIR-----DALAQALK 435

RESULT 17
US-10-032-585-7120
Sequence 7120, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Charles, Boone
APPLICANT: Howard, Bussey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-003-999
CURRENT APPLICATION NUMBER: US/10/032,585
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7120
LENGTH: 1259
TYPE: PRT
ORGANISM: Candida albicans
US-10-032-585-7120

Query Match 6.6%; Score 99.5; DB 12; Length 1259;
Best Local Similarity 21.8%; Pred. No. 5.7;
Matches 65; Conservative 59; Mismatches 97; Indels 77; Gaps 16;
QY 13 VKSAIETADGALDLYNKYLQVDPWKTDETIKELSRKQBSQASVILVDIKVLLMDS 72
DB 358 VKSAI--AEQAILAFFKKNPEVL-----ETIKIDESYWKNLTE-----KAFLM-- 399
QY 73 QDKFEATQTVVWCGVVTOLLSAYILLFDEYNEKKAQAKDILIRILDGVKKLNEAQK 132

DB 400 -----RTFYQCN-ENQJHALMDANFPPELLDLSITLEKYSVRL-----KTINE-NE 444
QY 133 SLTSSQSFNNASGKLLALDLSQTLNDFSEKSSVFSQSDVRKEA--YAGAAAGIVAGPF 190
DB 445 NLIKTWETYN--AKIDELNQI---FS-----LENQISRINTDADNFRKLSNIEED-- 491
QY 191 GLIISYTAAGVIEGKLIPELNNELKTVQNFPT-----SLSATVQKANK 234
DB 492 --IIEINIAKDLFX-KRIKQKNNSGNLEDLITEENQEIADQIKDFMEDLCQQLDINK 548
QY 235 DIDAQKLKATEIAAIGEIKTETTRFYVDYDMLSLKGAAKMINTCNEYQQRH 292
DB 549 NLDEIEHQPEDITAKLELQTK-----YDS-CIRALEPTSELXIQTVQIFEQEH 596

RESULT 18
US-10-128-714-3039
Sequence 3039, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wengqi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroshkin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3039
LENGTH: 1095
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-10-128-714-3039

Query Match 6.5%; Score 98; DB 15; Length 1095;
Best Local Similarity 19.4%; Pred. No. 6.4;
Matches 62; Conservative 56; Mismatches 92; Indels 110; Gaps 13;
QY 7 EQTVFVKSATETADGALDLYNKYLQVDPWKTDETIKELSRKQBSQASV--LVGD 64
DB 284 EKDIIEATNALVPVDEKVDITRKKVER-----FASRIAEIG--KERDQAANVQKLEKD 335
QY 65 IKVLLMDSQDKYFEATQTVVWCGVVTOLLSAYILLFDEYNEKKAQAKDILIRILDGV 124
DB 336 LKVV-----EKAQAQWEA----- 348
QY 125 KCLNEAQKSLTSSQSFNNASGKLLALDSQ-----LTNDFSEKSSVFSQSDVRKEAVAG 180
DB 349 ----EWHKTM-----SNKGGQLSESDQYKMKKEVSKRSEAEQINLNLKQRKTE 397
QY 181 AAGIVAGPGLIISYTAAGVIEGK--LIPELNNKLTQVNFFTSISATVQKANKIDA 238
DB 398 AEA-----YNSLKSFKPOSTEWQLKSVENDQTILTERKSALNDTVTKTSKELDR 445
QY 239 AKLK--LATEIAAIGEIKTETE-----TTRFYVDYD-----LMLSLLK--- 275
DB 446 KKKELNALTSELRISQMETELEBEKVQVVLKLLLEADGGKKQTERELRAKELISTLKRIF 505

QY 276 -CAAKKMINTCNEYQORHCK 294
Db 506 PGVKGVRSDLCRPHKQKVAE 525

RESULT 19

US-09-964-256A-2
; Sequence 2, Application US/09964256A
; Publication No. US20030049727A1
; GENERAL INFORMATION:
; APPLICANT: CURTIS, RORY A. J.
; TITLE OF INVENTION: 2568, A NOVEL HUMAN CALCIUM CHANNEL
; FILE REFERENCE: MNI-194
; CURRENT APPLICATION NUMBER: US/09/964,256A
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/235,018
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1223
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-964-256A-2

Query Match 6.5%; Score 98; DB 11; Length 1223;
Best Local Similarity 23.9%; Pred. No. 7.5;
Matches 61; Conservative 38; Mismatches 92; Indels 64; Gaps 12;

QY 10 VEVVKSATETADGALDLYNKY-----LDQVTPWKTFDETIKELSRFKQYEQ 56
Db 193 LQIAKDAQVILSAIDEDHKISVLTADTVTRCSDQC--YKTF-----LSPATSETKR 244
QY 57 EASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVTTQLLSAYILLFDEYNEKKAQAQKIL 116
Db 245 KMSTFVSSVK--SSDSPTQHAVGQ-----KAFQIRSTNNNTKFOANTDMV 289
QY 117 IRLDDGV--KKLNEAQK-----SLTSSQSFNNASGKLLALDSQLTDFSEKSSYFQSOV 170
Db 290 IYLSAGITSKDSSEEDKATLQVINEENSFLNNS--VMILTYALMND-----GV 337
QY 171 DRIRKEAYAGAAAGVAGPFGLLISYIAAGVIEGKL--IPELNNRLKTVQNFTTSLSAT 228
Db 338 TGLKELAFRLDLAEQNSGYG--VPDRTPALPVKSGMMVNLQSLNLETTVGRFYTNL--- 392
QY 229 VKQANKDIDAAXKL 243
Db 393 ---PNRMIDEAVFSL 404

RESULT 20

US-10-146-733-44
; Sequence 44, Application US/10146733
; Publication No. US20030165891A1
; GENERAL INFORMATION:
; APPLICANT: CURTIS, RORY A.J.
; APPLICANT: GLUCKSMANN, MARIA A.
; TITLE OF INVENTION: NOVEL TWIK-6, TWIK-7, IC3927, TWIK-8, IC47611, IC47615,
; TITLE OF INVENTION: HXMDA-1, TWIK-9, alpha2delta-4, 54414, AND 53763 MOLECULES
; FILE REFERENCE: MNI-248
; CURRENT APPLICATION NUMBER: US/10/146,733
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/185,938
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/515,520
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 09/518,866
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: US 60/195,734

; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: US 60/195,993
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/199,799
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 60/233,537
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/235,018
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 60/235,059
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US 60/256,240
; PRIOR FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 60/256,588
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 60/258,028
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 09/796,720
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/828,035
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 09/833,081
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: US 09/843,128
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 09/957,683
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 09/964,252
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 09/964,256
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 10/024,623
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 44
; LENGTH: 1223
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-146-733-44

Query Match 6.5%; Score 98; DB 12; Length 1223;
Best Local Similarity 23.9%; Pred. No. 7.5;
Matches 61; Conservative 38; Mismatches 92; Indels 64; Gaps 12;

QY 10 VEVVKSATETADGALDLYNKY-----LDQVTPWKTFDETIKELSRFKQYEQ 56
Db 193 LQIAKDAQVILSAIDEDHKISVLTADTVTRCSDQC--YKTF-----LSPATSETKR 244
QY 57 EASVLVGDIKVLLMDSQDKYFEATQTVYEWCGVTTQLLSAYILLFDEYNEKKAQAQKIL 116
Db 245 KMSTFVSSVK--SSDSPTQHAVGQ-----KAFQIRSTNNNTKFOANTDMV 289
QY 117 IRLDDGV--KKLNEAQK-----SLTSSQSFNNASGKLLALDSQLTDFSEKSSYFQSOV 170
Db 290 IYLSAGITSKDSSEEDKATLQVINEENSFLNNS--VMILTYALMND-----GV 337
QY 171 DRIRKEAYAGAAAGVAGPFGLLISYIAAGVIEGKL--IPELNNRLKTVQNFTTSLSAT 228
Db 338 TGLKELAFRLDLAEQNSGYG--VPDRTPALPVKSGMMVNLQSLNLETTVGRFYTNL--- 392
QY 229 VKQANKDIDAAXKL 243
Db 393 ---PNRMIDEAVFSL 404

RESULT 21

US-10-128-714-8039
; Sequence 8039, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: JIANG, BO
; APPLICANT: HU, WENQI

```
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8039
; LENGTH: 1277
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-8039

Query Match
; Sequence 566, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5060
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-5060

Query Match
; Sequence 97,5; DB 9; Length 759;
; Best Local Similarity 21.8%; Pred. No. 4.2;
; Matches 57; Conservative 48; Mismatches 94; Indels 63; Gaps 10;

QY 33 QVVPKTFDETIKLSRFKQY-----SQEASVL-----VGDIKVLLMDSQDK 75
Db 202 EVVPDQVDDIDAEIALFKQALEGVADMRALSSKLASQLRKEBRALFDVYVLLMLDASI 261

; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8039
; LENGTH: 1277
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-8039

Query Match
; Sequence 566, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5060
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-5060

Query Match
; Sequence 97,5; DB 9; Length 759;
; Best Local Similarity 21.8%; Pred. No. 4.2;
; Matches 57; Conservative 48; Mismatches 94; Indels 63; Gaps 10;

QY 33 QVVPKTFDETIKLSRFKQY-----SQEASVL-----VGDIKVLLMDSQDK 75
Db 202 EVVPDQVDDIDAEIALFKQALEGVADMRALSSKLASQLRKEBRALFDVYVLLMLDASI 261
```



```
; SEQ ID NO 2
; LENGTH: 1945
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-2

Query Match
Best Local Similarity 6.4%; Score 97.5; DB 11; Length 1945;
Matches 66; Conservative 51; Mismatches 108; Indels 89; Gaps 13;

QY 7 EQTVVVKSAIETADGA-----LDLYNKYLDQVVPKTFDETIKELSRF-----K 51
Db 1173 EQEVTVLKALDEETRSHEAQVQEMRQKHAQAV-----BELTEQLSQFKRAKANDK 1226
QY 52 QEYSQASVLVDIKVILMDSQDKYFEATQVYEWCGVWVTLQSSAVILLDFEYNEKKA 111
Db 1227 QTLEKENADLAGELRVL-----GQAKQEV-----EKKKKLEA 1259
QY 112 QKDILIRILDGVK-----KLENAQKSLTSSQSFNNASGKLLALDSQLTNDFSEK 164
Db 1260 QVQELQSKCSGGERARAEINLNDKVHKLQNEVESVTGMLNEAEGKAI-----K 1315
QY 165 YFQSQVDRIKAEYAGAAGIVAGPFGLLIISYSIAAGVIE---GKLIPELNRLKTVQNF 221
Db 1362 ---LERHISTLNQLSDSKKKLQDPFASTVEALEEGK-----KRFQKEIENLTQOYBEKAA 1413
QY 279 --KKMINTCNEYQQ 290
Db 1414 AYDKLEKTKNRLQQ 1427

RESULT 27
US-10-341-434-103
; Sequence 103, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR FILING DATE: 2003-07-18
; PRIOR FILING DATE: 2002-01-15
; PRIOR FILING DATE: 2002-01-15
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 103
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-103

Query Match
Best Local Similarity 6.4%; Score 97.5; DB 12; Length 1972;
Matches 66; Conservative 51; Mismatches 108; Indels 89; Gaps 13;

QY 7 EQTVVVKSAIETADGA-----LDLYNKYLDQVVPKTFDETIKELSRF-----K 51
Db 1173 EQEVTVLKALDEETRSHEAQVQEMRQKHAQAV-----BELTEQLSQFKRAKANDK 1226
QY 52 QEYSQASVLVDIKVILMDSQDKYFEATQVYEWCGVWVTLQSSAVILLDFEYNEKKA 111
Db 1227 QTLEKENADLAGELRVL-----GQAKQEV-----EKKKKLEA 1259
QY 112 QKDILIRILDGVK-----KLENAQKSLTSSQSFNNASGKLLALDSQLTNDFSEK 164
Db 1260 QVQELQSKCSGGERARAEINLNDKVHKLQNEVESVTGMLNEAEGKAI-----K 1315
QY 165 YFQSQVDRIKAEYAGAAGIVAGPFGLLIISYSIAAGVIE---GKLIPELNRLKTVQNF 221
Db 1362 ---LERHISTLNQLSDSKKKLQDPFASTVEALEEGK-----KRFQKEIENLTQOYBEKAA 1413
QY 279 --KKMINTCNEYQQ 290
Db 1421 AYDKLEKTKNRLQQ 1434

RESULT 28
US-10-171-311-162
; Sequence 162, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersht, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-06-13
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-162

Query Match
Best Local Similarity 6.4%; Score 97.5; DB 15; Length 1972;
Matches 66; Conservative 51; Mismatches 108; Indels 89; Gaps 13;

QY 7 EQTVVVKSAIETADGA-----LDLYNKYLDQVVPKTFDETIKELSRF-----K 51
Db 1173 EQEVTVLKALDEETRSHEAQVQEMRQKHAQAV-----BELTEQLSQFKRAKANDK 1226
QY 52 QEYSQASVLVDIKVILMDSQDKYFEATQVYEWCGVWVTLQSSAVILLDFEYNEKKA 111
Db 1227 QTLEKENADLAGELRVL-----GQAKQEV-----EKKKKLEA 1259
QY 112 QKDILIRILDGVK-----KLENAQKSLTSSQSFNNASGKLLALDSQLTNDFSEK 164
Db 1260 QVQELQSKCSGGERARAEINLNDKVHKLQNEVESVTGMLNEAEGKAI-----K 1315
QY 165 YFQSQVDRIKAEYAGAAGIVAGPFGLLIISYSIAAGVIE---GKLIPELNRLKTVQNF 221
Db 1362 ---LERHISTLNQLSDSKKKLQDPFASTVEALEEGK-----KRFQKEIENLTQOYBEKAA 1413
QY 279 --KKMINTCNEYQQ 290
Db 1414 AYDKLEKTKNRLQQ 1427
```

```

Query Match      6.4%; Score 97.5; DB 11; Length 1979;
Best Local Similarity 21.0%; Pred. No. 16;
Matches 56; Conservative 51; Mismatches 108; Indels 89; Gaps 13;

QY 7 EQTEVVKSAIETADGA-----LDLYNKYLDQVTPKTFDETIKLSRP-----K 51
Db 1180 EQETVVKALDETRSHSAQVQEMRQKHAQAV-----EELTEQLEQFKAKANLDKNK 1233
QY 52 QEYSQASVVLGDIKVLMLDSQDKYFEATCTVYEGGVTTQLLSAVILLFDYNEKKASA 111
Db 1234 QYLEKNADLAGELRLV-----GQAKQEV-----BHKKKLEA 1266
QY 112 QKDILRIIDDGVK-----KXNEAKSLTSSOSFNNASGKLLALDSQITNDFSEKSS 164
Db 1267 QVQELQSGSDGERARAEIENKVKHLQNEVESVTGMLNEAEGKAI-----KLAQDVASLS 1322
QY 165 YFQSQVDRIRKAYAGAAAGIAGVAGPGLIISYIAAGVIE---GKLIPELNNRLKTVQNF 221
Db 1323 QLODTQELIQETROK-----LNVSTKLRLQEBERNLSQQLDDEMEAKQN- 1368
QY 222 FTLSATVVKQAKNDIAAKLKL---ATEIAAIGEIKTETETTRFYVDYDDMLSLKGA 278
Db 1369 ---LERHISTNLQISDKKKLQDPASTVEALEEGK-----KRFQKEIENLTQOVEKAA 1420
QY 279 ---KMINTCNEYQO 290
Db 1421 AYDKLEKIKNRLLQO 1434

RESULT 30
US-10-172-502-2
; Sequence 2, Application US/10172502
; Publication No. US20030185633A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES.
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 2189
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-2

Query Match      6.4%; Score 97.5; DB 11; Length 1979;
Best Local Similarity 21.0%; Pred. No. 16;
Matches 56; Conservative 51; Mismatches 108; Indels 89; Gaps 13;

QY 7 EQTEVVKSAIETADGA-----LDLYNKYLDQVTPKTFDETIKLSRP-----K 51
Db 1180 EQETVVKALDETRSHSAQVQEMRQKHAQAV-----EELTEQLEQFKAKANLDKNK 1233
QY 52 QEYSQASVVLGDIKVLMLDSQDKYFEATCTVYEGGVTTQLLSAVILLFDYNEKKASA 111
Db 1234 QYLEKNADLAGELRLV-----GQAKQEV-----BHKKKLEA 1266
QY 112 QKDILRIIDDGVK-----KXNEAKSLTSSOSFNNASGKLLALDSQITNDFSEKSS 164
Db 1267 QVQELQSGSDGERARAEIENKVKHLQNEVESVTGMLNEAEGKAI-----KLAQDVASLS 1322
QY 165 YFQSQVDRIRKAYAGAAAGIAGVAGPGLIISYIAAGVIE---GKLIPELNNRLKTVQNF 221
Db 1323 QLODTQELIQETROK-----LNVSTKLRLQEBERNLSQQLDDEMEAKQN- 1368
QY 222 FTLSATVVKQAKNDIAAKLKL---ATEIAAIGEIKTETETTRFYVDYDDMLSLKGA 278
Db 1369 ---LERHISTNLQISDKKKLQDPASTVEALEEGK-----KRFQKEIENLTQOVEKAA 1420
QY 279 ---KMINTCNEYQO 290
Db 1421 AYDKLEKIKNRLLQO 1434

RESULT 31
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2

Query Match      6.4%; Score 97.5; DB 10; Length 2285;
Best Local Similarity 19.9%; Pred. No. 20;
Matches 65; Conservative 58; Mismatches 106; Indels 97; Gaps 16;

QY 9 TVEVVKSAIETADGAL-----DLYNKYLDQVTPKTFDETIKLSRQKQVSOE 57
Db 1091 SVDDIKAAIKEMSDAMQFDSVQVLDGDIPTNTKQVAP---LNDLEKMAEKSISANE 1147
QY 58 ASVLVGDIKVLMLDSQDKYFEATCTVYEGGVV---TQLLSAYILLPDEYNE-----K 107
Db 1148 ANTLI-----QKDK--ELAQAISIENGVVKINRDEVIKQKRVKLDAYNDMTVYSNK 1196
QY 108 KASAKDILIRILD-----DGVKKLEAKSLLTSSQ-----SFNASGKLLALDSQ 154
Db 1197 LKTEVNAIKTLNADTLRIDSUKRKERKLDMSAEILSEVKSINNVADEKEL-KK 1255

Query Match      6.4%; Score 97.5; DB 12; Length 2189;
Best Local Similarity 18.7%; Pred. No. 19;
Matches 69; Conservative 59; Mismatches 136; Indels 105; Gaps 12;

QY 6 AEQTEVVKSAIETADGALDLYNKYLDQVTPKTFDETIKLSRQKQVSOEASVLVGD 65
Db 1827 ATKINLIKANNEATD---EEQNAIVQV-----EKELIKAKQOIA--GAVTNADV 1872
QY 66 KVLMLDSQDKYFE-----ATQTYEVCVVTTQLLSA 96
Db 1873 AYLLHDKCKNEIRETEPVINKKATAREBQTLTFNDKQKQAEANVQATVEERNSILAQLONI 1932
QY 97 YILLFDEYNEKKASAQXD-----LIRILD-----DGVKKLEAKSLLTSSQSFN 142
Db 1933 YDTAIGQIDQDRSNAQYDKTATLNLQTIHDLVHPKPKDAERTINDLARVTHLVQNYR 1992
QY 143 NASGK-----LLALDSQITNDFSEKSSYFOSQVDRIRKAYAGAAAGIAGVAGPGLII 194
Db 1993 KVSDRNKADALKAITALKLQMDREL--KTARTWADVAVLKR-----2032
QY 195 SYSTAGVIEGKIPELNNRLKTVQNFSTLSATVVKQAKNDIAAKL-----242
Db 2033 -FNVALGDIEA-VITEKENSLLRIDNIAQOTYAKFKALATPEQLAKYKALIDQVADGNR 2090
QY 243 LATEIRAIGEIKTETETTRFYVDYDDMLSLKGA-----KMINTCNEYQORHGKKT 296
Db 2091 MVBEDATLNDIKDQTOLIIDEILAIKLPAAVIKASPKVGQAPKVCPTPIKEDKQEVK 2150
QY 297 LFEVDPVAS 305
Db 2151 VKELPNTGS 2159
```

QY 155 LTNDFSEKSYFQSOVDRIK-----EAYAGAAGIVAGPFGGLIISYSIAAGVIEKLIIP 209
Db 1256 LEEXOLQPGGYSNSQISAMQSVKSAUESYISASE-----EATSTQ 1295
QY 210 ELNNRLKTVQNFFTSLSATVKQANKDIDAAKLKLAIEAIGIKETETETTFYVD-YDD 268
Db 1296 EMNK--QALVEAGSLENWIDQDEK-----ANE-----ETKSMYVVDKYKE 1335
QY 269 LMLSLKGAAK--KXINTCNEYQQRH 292
Db 1336 ALEKVAEIDKYNQVNDYKYSQKY 1361
RESULT 32
US-10-080-608A-66
; Sequence 66, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; NUMBER OF SEQ ID NOS: 180
; CURRENT FILING DATE: 2002-02-21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-080-608A-66

Query Match 6.4%; Score 97; DB 12; Length 961;
Best Local Similarity 22.3%; Pred. No. 6.5;
Matches 67; Conservative 52; Mismatches 125; Indels 56; Gaps 14;
QY 10 VEVVKSALFETADGALDLYNKYLDQVIPWKTFDETIKLSRFKQYSEASVVLGDIKVL 69
Db 666 LEELKQOIIST-----LKQNEQLQTAV-----TQVVSQIQHQKQYNL-----LKVQL 708
QY 70 -MDSQDK--YFEATQTVYEWGCVVTVLLSAVILLDFEYNEKKAS-----AQKDILIR 118
Db 709 GKDSQHOGPYTDGAQ----NNGVQPEBISR--LREIEELKSNRELQSLAEKDSLIE 761
QY 119 ILDDGVKK--LNEAQKSLTSSQSFNNASGKLALDLSQTNDFSEKSYFQSOVDRIK 176
Db 762 NLKSSQSPGTNEQSATAGDSEQIAELKQELATLKSQ--NSQSVEITKLQTEKQELLQK 820
QY 177 AVAGAAGIVAGPFGGLIISYSIAAGVIEGKL-----IPELNNRLKTVQNFFTSLSATVK 230
Db 821 TEAFKSAFVPGSEETVIATKTTD--VEGRISALLQETKELKNEIKALSEERTAIKEQLD 878
QY 231 QANKDIDAAKLKLAIEAIGIKETETTFYVDYDDMLSLKGAAKKXINTCNEYQ 290
Db 879 SSNETI-----AILQNEKNKLEVDITDSKKEQDQLLV--LLADQDQKIFSLKKNLKE 928
RESULT 33
US-10-370-685-155
; Sequence 155, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF-P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 155
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Bos taurus
US-10-370-685-155

Query Match 6.4%; Score 97; DB 12; Length 961;
Best Local Similarity 22.3%; Pred. No. 6.5;
Matches 67; Conservative 52; Mismatches 125; Indels 56; Gaps 14;
QY 10 VEVVKSALFETADGALDLYNKYLDQVIPWKTFDETIKLSRFKQYSEASVVLGDIKVL 69
Db 666 LEELKQOIIST-----LKQNEQLQTAV-----TQVVSQIQHQKQYNL-----LKVQL 708
QY 70 -MDSQDK--YFEATQTVYEWGCVVTVLLSAVILLDFEYNEKKAS-----AQKDILIR 118
Db 709 GKDSQHOGPYTDGAQ----NNGVQPEBISR--LREIEELKSNRELQSLAEKDSLIE 761
QY 119 ILDDGVKK--LNEAQKSLTSSQSFNNASGKLALDLSQTNDFSEKSYFQSOVDRIK 176
Db 762 NLKSSQSPGTNEQSATAGDSEQIAELKQELATLKSQ--NSQSVEITKLQTEKQELLQK 820
QY 177 AVAGAAGIVAGPFGGLIISYSIAAGVIEGKL-----IPELNNRLKTVQNFFTSLSATVK 230
Db 821 TEAFKSAFVPGSEETVIATKTTD--VEGRISALLQETKELKNEIKALSEERTAIKEQLD 878
QY 231 QANKDIDAAKLKLAIEAIGIKETETTFYVDYDDMLSLKGAAKKXINTCNEYQ 290
Db 879 SSNETI-----AILQNEKNKLEVDITDSKKEQDQLLV--LLADQDQKIFSLKKNLKE 928

RESULT 34
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.C11A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996
Query Match 6.4%; Score 97; DB 9; Length 6281;

; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 109
; LENGTH: 3562
; TYPE: PRP
; ORGANISM: Homo sapiens
US-10-341-434-109

Query Match 6.4%; Score 96.5; DB 12; Length 3562;
Best Local Similarity 17.9%; Pred. No. 47;
Matches 61; Conservative 66; Mismatches 136; Indels 77; Gaps 13;
QY 7 EQTVVKSALETADGALLDLYNKYLQVDPKTFDE---TIKELSRFKQYSQ-EASV-- 60
Db 1034 EKTYSLOKDLSEMHENMTQAEELVLRDFEYKTPDELQKAVEEMKRAKEAQKAKVKL 1093
QY 61 -----LVGDIKULLMDSQKYEATCTVVEW-----CGVTQLLSAYILLFDEYN 105
Db 1094 LTESVNSVIAQAPFPAQALAKKELETLTNYQMLCTRLNGCKCTLEEVWACWHELL-SYL 1152
QY 106 EKKASAKDILIRI-----LDGKVKLNEAKSLLTSQSFNASGKLLALDSOLTND-- 158
Db 1153 EKANKWLNVEBFKLTENIFGGAEEISEVLDSLENLMRSHEDNPNQIRILAQTLTGGV 1212
QY 159 ----FSEKSSYFQSDVRIRKEAYAGAAAGIVAGPFGLLIISYSTAAAGVIEGKLIPELNNR 214
Db 1213 MDLINEELETFSNRWRELBEEAVRRQK-----LRFQSIQSA-----QETESK 1255
QY 215 LKTVQNFSTLSATVQKANK-IDAALKLATEIAAIGETETETTFRVVDYDDL---- 269
Db 1256 LHLIQESLTFIDKQLAAVIAKVDAAQNPQEAQ-----KIQSDLISHISEEMKKN 1308
QY 270 -----MLSLKGAAKKMINTCNEYQOHRGKKTLPFVP 301
Db 1309 QKKEAQRVLSQIDVAQKQLQDVSMKFR-----LFQKP 1341

RESULT 38
US-09-976-782-6
; Sequence 6, Application US/09976782
; Publication No. US20030190715A1
; GENERAL INFORMATION:
; APPLICANT: Grosse et al
; TITLE OF INVENTION: No. US20030190715A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-157
; CURRENT APPLICATION NUMBER: US/09/976,782
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,113
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: 60/240,662
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/240,732
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/240,625
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/240,703
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/241,190
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/240,637
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/240,669
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/262,455
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/240,648
; PRIOR FILING DATE: 2000-10-16

; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 6
; LENGTH: 523
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-976-782-6

Query Match 6.3%; Score 96; DB 12; Length 523;
Best Local Similarity 18.8%; Pred. No. 3.4;
Matches 59; Conservative 55; Mismatches 99; Indels 100; Gaps 12;
QY 24 LDLYNKYLQDVIPWK-TFDETIKEL-----SRFKQYSQSEASVLVDGIKVLMLDSQKYFE 78
Db 194 INLYSLDNLILGERGLDSELKNMEDLVEDFKKY-----EDEINK 235
QY 79 ATQTVYWGCVVTVQLLSAYILLFDEYNKKAQKDIILIRILDGKVKLNEAKSLLTSS 138
Db 236 RTAAENEFVTLKKOVD SAYM-----NKVELQAKVDALIDEI-DPLRTLYDAELSQQVTH 288
QY 139 QSFNNSGKLLALDSOLTNDSEKSSYFQSDVRIRKEAYAGAAA-----GIVAGP 189
Db 289 VENTV---VLSMDNNRLDLSIIAEYKAYVELIAQSRABAEAWYQTKVEELQVTAGK 345
QY 190 FGLIISYSTAAAGVIEGKLIPELNNRLKTVQNFSTLSATVQKANKIDDAK-----LKLAT 245
Db 346 HG-----DNLRTKNEIAELTRIQRLQGEADAQKQCCQLQT 384
QY 246 ETA-----AIGEI-----KTETETTFRVVDYDDL-----LS 272
Db 385 AIAABEORGEALAKDAQKIGLDLVAHQAKEDTLRLRDYQELMNVKALDVEIATYRK 444
QY 273 LKGAAKKVMINTC 285
Db 445 LLESQESRMSGEC 457

RESULT 39
US-10-080-608A-19
; Sequence 19, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 975
; TYPE: PRP
; ORGANISM: Drosophila melanogaster
US-10-080-608A-19

Query Match 6.3%; Score 96; DB 12; Length 975;
Best Local Similarity 21.1%; Pred. No. 8.2;
Matches 63; Conservative 56; Mismatches 118; Indels 62; Gaps 14;
QY 37 WKTFTETIKELSRFKQYSQSEASVLVDGIKVLMLDSQKYFEATQTVYEWCGVVTQL--- 93
Db 376 WRA-GETVKAEEQINMEDLMEAS-----TPNLEVEAAQTAATAAALAAQRTAL 422
QY 94 --LSAYILLFDEYNK-KASAKDILIRILDGKVKLNEAKSLLTSSQSFNNSGKLLA 150
Db 423 ANMSASVAV-----NEQRLATECERLYQQLDDKDEINO-----QSQYAEQLKEQWME 471
QY 151 LDSOLTNDSEKSSYFQSDVRIRKEAYAG--AAAGIVAGPFGLLIISYSTAAAGVIEGKLI- 207
Db 472 QEBLIANARREYET-LQSEWARIQENESAKBEVKEVLQALBELAVNYVDQKSEIDNNK 530

Db 588 SRGVKLLALLDEGGSCNPLLOAKAGLNGAVSLLRSQQPASBPQRILQAGNCGSA 647
QY 145 SGKLLALDSQNTNPFSEKSYFQSQVDRIRKEAVAGAAAGIVAGPFGLLISY-SIA---- 199
Db 648 SGEILL---QSITE--SDTDPHFQVALMQLAKAVMSAAA-----LVLYKYSVAQRTE 694
QY 200 -AGVTEGLKILPELNNRLKTVQNFETS-----LSATVKQANKDIDAAKL--KLATEIAAI 250
Db 695 DAGLTQVIAADTQCALSTSEVACTKVVAFFISPVQC--EQGVEAGRLVAKHVEGCVSA 753
QY 251 GBIKTETETTFYVDYDDMLSLKGAAGKMTNCNEYQQ 290
Db 754 SQIATE-----DQQLRKVGAAATAVTALNELQ 783

RESULT 43

US-10-013-477-18
; Sequence 18, Application US/10013477
; Publication No. US20030049732A1

GENERAL INFORMATION:

; APPLICANT: Ni et al.

; TITLE OF INVENTION: Apoptosis Related Polynucleotides, Polypeptides, and Antibodies

; FILE REFERENCE: PT002PI

; CURRENT APPLICATION NUMBER: US/10/013,477

; PRIOR FILING DATE: 2001-12-13

; PRIOR APPLICATION NUMBER: 09/669,445

; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: PCT/US00/06642

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: 60/126,018

; PRIOR FILING DATE: 1999-03-24

; PRIOR APPLICATION NUMBER: 60/139,638

; PRIOR FILING DATE: 1999-06-17

; PRIOR APPLICATION NUMBER: 60/149,449

; PRIOR FILING DATE: 1999-08-18

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 18

; LENGTH: 613

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-013-477-18

Query Match 6.3%; Score 95; DB 15; Length 613;

Best Local Similarity 21.1%; Pred. No. 5.2;

Matches 59; Conservative 48; Mismatches 88; Indels 84; Gaps 13;

QY 37 WKTFTETIKELSRFKQEYSQEAS-----VLVGDIKVLLMDS 72
Db 180 WFSDDPNVTKTLRFQWNGKERSIVFPSPFVVSADLPHIENGGAVALTGK-KVQLDV 238
QY 73 QD---KYFATQTVIEWCGVVT-----QLLSAIVLLFDEYNEKASQAQ---KDILIRLDD 122
Db 239 RDNWKLNDGSOITVEKCLIAATGTPRSLSAI-----DRAGAIVKSRRTTLFRKIGD 289
QY 123 --GVKKLINEAQSLTSSQSFNNA-----SGKLLALDSQLTNDSEKSS-----YFQ 167
Db 290 FRSLKISREVKSITIIIGGFLGSELACALGRKARALGTEVIQLPFGKNGMKILPEYLS 349
QY 168 S-QVDRIRKEAVAGAAAGIVAGPFGLLISYSTAAGVIEGKLIPELNNRLKTVQNFETSLS 226
Db 350 NWTMEKVRRE-----GVKVPNFAIVQS-----VGVSCKLILKLDGRKV----- 389
QY 227 ATVKQANKDIDAAKLKLATEIAAIGIKTETETTFYVD 265
Db 390 ----ETDHIIVAAVGLPEPNVELAKTGGLGLEDSDFGGFRVN 424

RESULT 44

US-10-013-477-26
; Sequence 26, Application US/10013477
; Publication No. US20030049732A1

GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Apoptosis Related Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT002PI
; CURRENT APPLICATION NUMBER: US/10/013,477
; PRIOR FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: 09/669,445
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: PCT/US00/06642
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 60/126,018
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: 60/139,638
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: 60/149,449
; PRIOR FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 26
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-013-477-26

Query Match 6.3%; Score 95; DB 15; Length 613;

Best Local Similarity 21.1%; Pred. No. 5.2;

Matches 59; Conservative 48; Mismatches 88; Indels 84; Gaps 13;

QY 37 WKTFTETIKELSRFKQEYSQEAS-----VLVGDIKVLLMDS 72
Db 180 WFSDDPNVTKTLRFQWNGKERSIVFPSPFVVSADLPHIENGGAVALTGK-KVQLDV 238
QY 73 QD---KYFATQTVIEWCGVVT-----QLLSAIVLLFDEYNEKASQAQ---KDILIRLDD 122
Db 239 RDNWKLNDGSOITVEKCLIAATGTPRSLSAI-----DRAGAIVKSRRTTLFRKIGD 289
QY 123 --GVKKLINEAQSLTSSQSFNNA-----SGKLLALDSQLTNDSEKSS-----YFQ 167
Db 290 FRSLKISREVKSITIIIGGFLGSELACALGRKARALGTEVIQLPFGKNGMKILPEYLS 349
QY 168 S-QVDRIRKEAVAGAAAGIVAGPFGLLISYSTAAGVIEGKLIPELNNRLKTVQNFETSLS 226
Db 350 NWTMEKVRRE-----GVKVPNFAIVQS-----VGVSCKLILKLDGRKV----- 389
QY 227 ATVKQANKDIDAAKLKLATEIAAIGIKTETETTFYVD 265
Db 390 ----ETDHIIVAAVGLPEPNVELAKTGGLGLEDSDFGGFRVN 424

RESULT 45

US-10-207-655-204

; Sequence 204, Application US/10207655

; Publication No. US20030118592A1

GENERAL INFORMATION:

; APPLICANT: Ledbetter, Jeffrey A.

; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS

; FILE REFERENCE: 390069.401C1

; CURRENT APPLICATION NUMBER: US/10/207,655

; CURRENT FILING DATE: 2002-07-25

; NUMBER OF SEQ ID NOS: 426

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 204

; LENGTH: 613

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-207-655-204

Query Match 6.3%; Score 95; DB 15; Length 613;

Best Local Similarity 21.1%; Pred. No. 5.2;

Matches 59; Conservative 48; Mismatches 88; Indels 84; Gaps 13;

QY 37 WKTFTETIKELSRFKQEYSQEAS-----VLVGDIKVLLMDS 72

Db 180 WFSDDPNVTLRFKQNGKRSYFQPPSYVSAQDLPHIENGVAULTG-KVQLDV 238
Qy 73 QD---KYFEATQVWVWCVVT---QLLSAYILLFDEYNEKASQAQ---KDLIRILD 122
Db 239 RDNWKLNDGSOITYEKLJATGTGTPRSLSA-----DRAGAEVKSRTTLFRKIGD 289
Qy 123 --GVKLNKAQKLLTSSQSFNNA-----SGKLLALDSQLTNDFSEKSS-----YFQ 167
Db 290 FRSLKISREVKISITIGGFLGSELACALGRARALGTEVIOFPPEKNGMKILPEYLS 349
Qy 168 S-OVDRIRKAYAGAAAGIVAGPGLLIYSIAAGVIEGKLIPELNNRLKTVQNFPTSLS 226
Db 350 NWTMEKVR-----GVKWNVAIVQS-----GVSSGKLLIKLKGDKRV-----389
Qy 227 ATVKQANKOIDAALKLATEIAAIGIKTETETETRYVD 265
Db 390 ---ETDHIVAAGVLEPNVELAKTGGLLEIDSDFGFRVN 424

RESULT 46

US-09-756-071B-15

; Sequence 15, Application US/09756071B

; Patent No. US20020052307A1

; GENERAL INFORMATION:

; APPLICANT: Tryggevason, Karl

; Kallunki, Pekka

; Pyke, Charles

; TITLE OF INVENTION: Laminin Chains: Diagnostic Uses

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fay Sharpe Pagan Minnich & McKe

; STREET: 1100 Superior Ave, Suite 700

; CITY: Cleveland

; STATE: Ohio

; COUNTRY: USA

; ZIP: 44114

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/756,071B

; FILING DATE: 08-Jan-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 09/663,147

; FILING DATE: 150-September 2000

; ATTORNEY/AGENT INFORMATION:

; NAME: Minnich, Richard, J.

; REGISTRATION NUMBER: 24,175

; REFERENCE/DOCKET NUMBER: TRV 20014

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 216-861-5582

; TELEFAX: 216-241-1666

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1111 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 15:

US-09-756-071B-15

Query Match

Best Local Similarity 6.3%; Score 95; DB 9; Length 1111;

Matches 48; Conservative 51; Mismatches 97; Indels 72; Gaps 9;

Qy 10 VEVVKSALFADGALDLYNKYLQVLPWKTFDETIELSRFKQYISQ-----EAS 59

Db 868 VEEAKRIKQKADSLSSLVTRHMD-----FKRTQKNLGNWKBEAQQLLQNGKSGREKSD 921

Qy 60 VLVGDIKVLMDSQKYEATQTVYVWCVVWVLTQLLSAYILLFD-EYNEKASQAQ-----113
Db 922 QLLSRANLAKRAQEAALSMGNATFEYESILKNLRE-----FDLQVDRNRKAAEAEAMKRL 976
Qy 114 DILIRILDDGVYKLNKAQKSL---LTSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQV 170
Db 977 SYISCKVSDASDKTQQARALGSAADAQACRAKNGAGEALSEISSEI-----EQEI 1025
Qy 171 DRIRKEAYAGAAAGIVAGPGLLIYSIAAGVIEGKLIPELNNRLKTVQNFPTSLSATVK 230
Db 1026 GSLNLEAV-TADGALAMEKGL-----ASLKSEMR 1054
Qy 231 QANKDIDAADAKLATEIAAIGIKTETE 258
Db 1055 EVELERKELEFDTNMDAVQMVITEAQ 1082

RESULT 47

US-10-227-738-15

; Sequence 15, Application US/10227738

; Publication No. US20030100529A1

; GENERAL INFORMATION:

; APPLICANT: Tryggevason, Karl

; Kallunki, Pekka

; Pyke, Charles

; TITLE OF INVENTION: Laminin Chains: Diagnostic and

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

; STREET: 300 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/227,738

; FILING DATE: 26-Aug-2002

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/800,593

; FILING DATE: 18-FEB-1997

; APPLICATION NUMBER: US 08/317,450

; FILING DATE: 04-OCT-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Chao, Mark

; REGISTRATION NUMBER: 37,293

; REFERENCE/DOCKET NUMBER: 94,778-B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-913-0001

; TELEFAX: 312-913-0002

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1111 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 15:

US-10-227-738-15

Query Match

Best Local Similarity 6.3%; Score 95; DB 15; Length 1111;

Matches 48; Conservative 51; Mismatches 97; Indels 72; Gaps 9;

Qy 10 VEVVKSALFADGALDLYNKYLQVLPWKTFDETIELSRFKQYISQ-----EAS 59

Db 868 VEEAKRIKQKADSLSSLVTRHMD-----FKRTQKNLGNWKBEAQQLLQNGKSGREKSD 921

QY 60 VLVDGIKVLMDSDQKYFATQTVYWCWVTVLLSAYILLFD-EYNEKASAKQ-----113
 Db 922 QLLSRANLAKSRAQEAALSMGNATFYEVESILKNLRE-----FDLQVDRKRAEABEAMKRL 976
 QY 114 DILIRILDGDKVYKLNKAEQKSL---LTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSOV 170
 Db 977 SYISQKVSADSKTQQAERALGSAADAQRAKNGAGALEISSEI-----EQEI 1025
 QY 171 DRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVK 230
 Db 1026 GSLNLEAV-TADGALAMEKGL-----ASLKSEMR 1054
 QY 231 QANKDIDAAKLKLATEIAAIGEIKTETE 258
 Db 1055 EVEGELERKELEFDINDAVQWITEAQ 1082

RESULT 48

US-09-756-071B-13
 ; Sequence 13, Application US/09756071B
 ; Patent No. US20020052307A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tryggvason, Karl
 ; Kallunki, Pekka
 ; Pyke, Charles
 ; TITLE OF INVENTION: Laminin Chains: Diagnostic Uses
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fay Sharpe Fagan Minnich & McKee
 ; STREET: 1100 Superior Ave, Suite 700
 ; CITY: Cleveland
 ; STATE: Ohio
 ; COUNTRY: USA
 ; ZIP: 44114
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/756,071B
 ; FILING DATE: 08-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/663,147
 ; FILING DATE: 150-September 2000
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Minnich, Richard, J.
 ; REGISTRATION NUMBER: 24,175
 ; REFERENCE/DOCKET NUMBER: TRV 20014
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 216-861-5582
 ; TELEFAX: 216-241-1666
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1193 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Query Match 6.3%; Score 95; DB 9; Length 1193;
 Best Local Similarity 17.9%; Pred. No. 14;
 Matches 48; Conservative 51; Mismatches 97; Indels 72; Gaps 9;

QY 10 VEVVKSATETAGDALDLYNKYLDQVTPKTFDTIKELSRFKQEYSQ-----EAS 59
 Db 868 VEEAKRIKQKADSLSLVTRHDE-----FKRTQKNLGNWKEEAQQLLNGKSGREKSD 921
 QY 60 VLVDGIKVLMDSDQKYFATQTVYWCWVTVLLSAYILLFD-EYNEKASAKQ-----113
 Db 922 QLLSRANLAKSRAQEAALSMGNATFYEVESILKNLRE-----FDLQVDRKRAEABEAMKRL 976

QY 114 DILIRILDGDKVYKLNKAEQKSL---LTSSQSFNNASGKLLALDSQLTNDSEKSSYFQSOV 170
 Db 977 SYISQKVSADSKTQQAERALGSAADAQRAKNGAGALEISSEI-----EQEI 1025
 QY 171 DRIRKEAYAGAAAGIVAGPGLIISYSIAAGVIEGKLIPELNNRLKTVQNFFTSLSATVK 230
 Db 1026 GSLNLEAV-TADGALAMEKGL-----ASLKSEMR 1054
 QY 231 QANKDIDAAKLKLATEIAAIGEIKTETE 258
 Db 1055 EVEGELERKELEFDINDAVQWITEAQ 1082

RESULT 49

US-10-227-738-13
 ; Sequence 13, Application US/10227738
 ; Publication No. US20030100529A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tryggvason, Karl
 ; Kallunki, Pekka
 ; Pyke, Charles
 ; TITLE OF INVENTION: Laminin Chains: Diagnostic and
 ; Therapeutic Use
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 ; STREET: 300 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/227,738
 ; FILING DATE: 26-Aug-2002
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/800,593
 ; FILING DATE: 18-FEB-1997
 ; APPLICATION NUMBER: US 08/317,450
 ; FILING DATE: 04-OCT-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Chao, Mark
 ; REGISTRATION NUMBER: 37,293
 ; REFERENCE/DOCKET NUMBER: 94,778-B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-913-0001
 ; TELEFAX: 312-913-0002
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1193 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Query Match 6.3%; Score 95; DB 15; Length 1193;
 Best Local Similarity 17.9%; Pred. No. 14;
 Matches 48; Conservative 51; Mismatches 97; Indels 72; Gaps 9;

QY 10 VEVVKSATETAGDALDLYNKYLDQVTPKTFDTIKELSRFKQEYSQ-----EAS 59
 Db 868 VEEAKRIKQKADSLSLVTRHDE-----FKRTQKNLGNWKEEAQQLLNGKSGREKSD 921
 QY 60 VLVDGIKVLMDSDQKYFATQTVYWCWVTVLLSAYILLFD-EYNEKASAKQ-----113
 Db 922 QLLSRANLAKSRAQEAALSMGNATFYEVESILKNLRE-----FDLQVDRKRAEABEAMKRL 976

